

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:27:25 ; Search time 82.85 Seconds
(without alignments) 57.648 Million cell updates/sec

Title: US-09-813-214A-1

Perfect score: 209

Sequence: 1 IGISEADGCGGAGNARDKS.....GDIAQALGSGSTAIQDKIV 43

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802:*
1: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 209 | 100.0 | 43 | 18 | AAW32274 |
| 2 | 209 | 100.0 | 43 | 22 | AAE0703 |
| 3 | 209 | 100.0 | 2123 | 22 | AAE00701 |
| 4 | 190 | 90.9 | 40 | 22 | AAE00705 |
| 5 | 121 | 57.9 | 24 | 18 | AAW32276 |
| 6 | 121 | 57.9 | 24 | 22 | AAE00702 |
| 7 | 90 | 43.1 | 1992 | 17 | AAW04505 |
| 8 | 90 | 43.1 | 1992 | 22 | AAE69133 |
| 9 | 90 | 43.1 | 1992 | 22 | AAE69137 |
| 10 | 90 | 43.1 | 2047 | 22 | AAE69134 |
| 11 | 90 | 43.1 | 2053 | 22 | AAE69135 |

| | | | | | |
|----|------|------|------|----|----------|
| 12 | 90 | 43.1 | 2314 | 22 | AAE69136 |
| 13 | 67 | 32.5 | 1778 | 22 | ABE52677 |
| 14 | 68 | 32.1 | 2042 | 19 | AAW56319 |
| 15 | 63 | 30.1 | 261 | 20 | AAV13486 |
| 16 | 61 | 29.2 | 2039 | 19 | AAW56332 |
| 17 | 60.5 | 28.9 | 608 | 22 | ABE64278 |
| 18 | 60 | 28.7 | 278 | 15 | AAE6490 |
| 19 | 60 | 28.7 | 278 | 15 | AAE6490 |
| 20 | 60 | 28.7 | 278 | 18 | AAW5025 |
| 21 | 60 | 28.7 | 278 | 22 | ABE62342 |
| 22 | 60 | 28.7 | 289 | 21 | ABE7262 |
| 23 | 60 | 28.7 | 289 | 21 | AAE40485 |
| 24 | 60 | 28.7 | 327 | 21 | AAE27261 |
| 25 | 60 | 28.7 | 449 | 22 | ABE6828 |
| 26 | 58.5 | 28.0 | 464 | 20 | AAE08400 |
| 27 | 56 | 26.8 | 591 | 22 | ABG18023 |
| 28 | 55.5 | 26.6 | 240 | 22 | AAE64046 |
| 29 | 55.5 | 26.6 | 919 | 22 | ABE29111 |
| 30 | 55 | 26.3 | 193 | 20 | AAV37714 |
| 31 | 55 | 26.3 | 252 | 21 | AAE19014 |
| 32 | 55 | 26.3 | 889 | 19 | AAE68208 |
| 33 | 55 | 26.3 | 1114 | 22 | ABG08489 |
| 34 | 55 | 26.3 | 1173 | 22 | ABG08134 |
| 35 | 54.5 | 26.1 | 166 | 21 | AAE27234 |
| 36 | 54.5 | 26.1 | 187 | 22 | ABE58038 |
| 37 | 54.5 | 26.1 | 467 | 21 | AAV51615 |
| 38 | 54 | 25.8 | 258 | 20 | AAV13485 |
| 39 | 54 | 25.8 | 456 | 22 | ABE64582 |
| 40 | 54 | 25.8 | 559 | 22 | AAE65573 |
| 41 | 53.5 | 25.6 | 700 | 22 | AAU35467 |
| 42 | 53.5 | 25.6 | 704 | 21 | AAE15891 |
| 43 | 53.5 | 25.6 | 704 | 22 | AAU34770 |
| 44 | 53.5 | 25.6 | 704 | 22 | AAU38372 |
| 45 | 53.5 | 25.6 | 1438 | 22 | ABG03631 |

ALIGNMENTS

| | | |
|----------|---|---------------------------|
| RESULT 1 | AAW32274 | standard; peptide; 43 AA. |
| XX | AAW32274: | |
| AC | 08-MAY-1998 | (first entry) |
| XX | | |
| DE | M. catarrhalis outer membrane protein (OMP)-106 peptide fragment 1. | |
| XX | | |
| KW | Outer membrane protein-106; OMP106; vaccine; immune response; | |
| KW | Cytotoxic antibody; Moraxella catarrhalis. | |
| OS | Moraxella catarrhalis. | |
| XX | | |
| XX | WO9741731-A1. | |
| PN | | |
| XX | 13-NOV-1997. | |
| PD | | |
| XX | 28-APR-1997; 97WO-US07679. | |
| PF | | |
| XX | 03-MAY-1996; 96US-0642712. | |
| PR | | |
| XX | (ANTE-) ANTEX BIOLOGICS INC. | |
| PA | Flotilla L, Tucker K; | |
| XX | WPI; 1997-558601/51. | |
| XX | Outer membrane protein, OMP106, of Moraxella catarrhalis - used in | |
| PT | vaccines for producing immune responses against M. catarrhalis | |
| XX | | |
| PS | Claim 9; page 23; 78pp; English. | |
| XX | | |

CC This is a peptide fragment of a novel outer membrane protein-106 (OMP106)
 CC polypeptide. The OMP106 is an outer membrane polypeptide of Moraxella
 CC catarhalis, an haemagglutinating cultivar. The peptide fragment can
 CC specifically bind to an antibody that binds the OMP106 polypeptide. The
 CC antibody is a cytotoxic antibody which mediates complement killing of
 CC M. catarhalis. The OMP106 polypeptide, and its peptide fragments can be
 CC used in vaccines and antigenic compositions. They can also be used for
 CC producing an immune response in an animal against M. catarhalis.
 XX
 SQ Sequence 43 AA;

Query Match 100.0%; Score 209; DB 18; Length 43;
 Best Local Similarity 100.0%; Pred. No. 1.1e-19;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dy 1 IGISEADGGKGANRDKSIAIGDIAQALGSGSIAIGDNKIV 43
 Db 1 IgiSeadggkgganargdksiaigdiagalsgsiaigdnkiv 43

RESULT 2

ID AAE00703 standard; peptide; 43 AA.

AC AAE00703;

DT 02-JUL-2001 (first entry)

DE N-terminal #1 of M. catarhalis outer membrane protein-106 (OMP106).

KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 KW bacterial infection; immunogen; cytotoxic; antibiotic;

OS Moraxella catarhalis.

PN US6214981-B1.

PD 10-APR-2001.

PF 12-NOV-1997; 97US-0968685.

PR 03-MAY-1996; 96US-0642712.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Tucker K, Plosila L, Tillman UF;

PS WPI: 2001-281002/29.

PT Novel nucleotide sequences encoding Moraxella catarhalis outer
 PT membrane protein-106 polypeptide, useful for diagnosis of bacterial
 PT infections and as vaccine against Moraxella catarhalis infection of
 PT mammals -

PS Claim 2; Column 31; 49pp; English.

CC The present sequence is N-terminal of haemagglutinating
 CC Moraxella catarhalis outer membrane protein-106 (OMP106). The OMP106 is
 CC used as a therapeutic and prophylactic vaccine against M. catarhalis
 CC infections of mammals. It is used for diagnosis of bacterial infections
 CC and as reagents for clinical or medical diagnosis of M. catarhalis
 CC infections and for scientific research on the properties of
 CC pathogenicity, virulence and infectivity of M. catarhalis. It is also
 CC used as a probe to identify other bacteria that encode a polypeptide
 CC specimens and to identify other bacteria that encode a polypeptide
 CC related to M. catarhalis OMP106. OMP106-derived polypeptides are used
 CC as ligands to detect antibodies elicited in response to M. catarhalis
 CC infections and also as immunogens for inducing M. catarhalis-specific
 CC antibodies which are useful in immunoassays to detect M. catarhalis in
 CC biological specimens. Cytotoxic antibodies are useful in passive
 CC immunisations against M. catarhalis.

XX
 SQ Sequence 43 AA;

Query Match 100.0%; Score 209; DB 22; Length 43;
 Best Local Similarity 100.0%; Pred. No. 1.1e-19;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dy 1 IGISEADGGKGANRDKSIAIGDIAQALGSGSIAIGDNKIV 43
 Db 1 IgiSeadggkgganargdksiaigdiagalsgsiaigdnkiv 43

RESULT 3

ID AAE00701 standard; Protein; 2123 AA.

AC AAE00701;

DT 02-JUL-2001 (first entry)

DE Moraxella catarhalis outer membrane protein-106 (OMP106).

KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 KW bacterial infection; immunogen; cytotoxic; antibiotic;

OS Moraxella catarhalis.

PN US6214981-B1.

PD 10-APR-2001.

PF 12-NOV-1997; 97US-0968685.

PR 03-MAY-1996; 96US-0642712.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Tucker K, Plosila L, Tillman UF;

PS WPI: 2001-281002/29.

PT Novel nucleotide sequences encoding Moraxella catarhalis outer
 PT membrane protein-106 polypeptide, useful for diagnosis of bacterial
 PT infections and as vaccine against Moraxella catarhalis infection of
 PT mammals -

PS Claim 7; Column 53-64; 49pp; English.

CC The present sequence is haemagglutinating Moraxella catarhalis outer
 CC membrane protein-106 (OMP106). The OMP106 is used as a therapeutic
 CC and prophylactic vaccine against M. catarhalis infections of mammals.
 CC It is used for diagnosis of bacterial infections and as reagents for
 CC clinical or medical diagnosis of M. catarhalis infections and for
 CC scientific research on the properties of pathogenicity, virulence and
 CC infectivity of M. catarhalis. It is also used as a probe to identify
 CC other bacteria that encode a polypeptide related to M. catarhalis
 CC OMP106. OMP106-derived polypeptides are used as ligands to detect
 CC antibodies elicited in response to M. catarhalis infections and also
 CC as immunogens for inducing M. catarhalis-specific antibodies which are
 CC useful in immunoassays to detect M. catarhalis in biological specimens.
 CC Cytotoxic antibodies are useful in passive immunisations against
 CC M. catarhalis.

SQ Sequence 2123 AA;

Query Match 100.0%; Score 209; DB 22; Length 2123;
 Best Local Similarity 100.0%; Pred. No. 9e-18;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IGISEADGKGGANARCKSAIGDIAQALGSOSIAIGDNKIV 43
 DB 69 igiseadgkkgganargdksiaigdiagalsgsiaigdnkiv 111

RESULT 4

ID AAE00705 standard; peptide: 40 AA.

AC AAE00705;

DT 02-JUL-2001 (first entry)

DE N-terminal #2 of M. catarrhalis outer membrane protein-106 (OMP106).

KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;

KM bacterial infection; immunogen; cytotoxic; antituberc;

OS Moraxella catarrhalis.

PN US6214981-B1.

PD 10-APR-2001.

PF 12-NOV-1997; 97US-0968685.

PR 03-MAY-1996; 96US-0642712.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Tucker K, Plosila L, Tillman WF;

DR WPI; 2001-281002/29.

PT Novel nucleotide sequences encoding Moraxella catarrhalis outer
 membrane protein-106 polypeptide, useful for diagnosis of bacterial
 infections and as vaccine against Moraxella catarrhalis infection of
 mammals -

PS Example: Column 31; 49pp: English.

CC The present sequence is N-terminal of haemagglutinating
 CC Moraxella catarrhalis outer membrane protein-106 (OMP106). The OMP106 is
 CC used as a therapeutic and prophylactic vaccine against M. catarrhalis
 CC infections of mammals. It is used for diagnosis of bacterial infections
 CC and as reagents for clinical or medical diagnosis of M. catarrhalis
 CC infections and for scientific research on the properties of
 CC pathogenicity, virulence and infectivity of M. catarrhalis. It is also
 CC used as a probe to identify the presence of M. catarrhalis in biological
 CC specimens and to identify other bacteria that encode a polypeptide
 CC related to M. catarrhalis OMP106. OMP106-derived polypeptides are used
 CC as ligands to detect antibodies elicited in response to M. catarrhalis
 CC infections and also as immunogens for inducing M. catarrhalis-specific
 CC antibodies which are useful in immunoassays to detect M. catarrhalis in
 CC biological specimens. Cytotoxic antibodies are useful in passive
 CC immunisations against M. catarrhalis.

CC Sequence 40 AA;

Query Match 90.9%; Score 190; DB 22; Length 40;
 Best Local Similarity 100.0%; Pred. No. 2.6e-17;
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IGISEADGKGGANARCKSAIGDIAQALGSOSIAIGD 39
 DB 2 igiseadgkkgganargdksiaigdiagalsgsiaigd 40

RESULT 5
 AAW32276

ID AAW32276 standard; peptide: 24 AA.

AC AAW32276;

DT 08-MAY-1998 (first entry)

DE M. catarrhalis outer membrane protein (OMP)-106 partial peptide fragment.

KW Outer membrane protein-106; OMP106; vaccine; immune response;

KM cytotoxic antibody; Moraxella catarrhalis; primer; probe.

OS Moraxella catarrhalis.

PN WO9741731-A1.

PD 13-NOV-1997.

PF 28-APR-1997; 97WO-US07679.

PR 03-MAY-1996; 96US-0642712.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Plosila L, Tucker K;

DR WPI; 1997-558601/51.

DR N-FSD; AAT86522.

PT Outer membrane protein, OMP106, of Moraxella catarrhalis - used in
 vaccines for producing immune responses against M. catarrhalis

PS Disclosure; Page 58; 78pp: English.

CC This is a partial sequence of an amino terminal peptide fragment of a
 CC novel outer membrane protein-106 (OMP106) polypeptide. The OMP106 is an
 CC outer membrane polypeptide of Moraxella catarrhalis, an haemagglutinating
 CC cultivar. The encoding DNA can be used as a 5' primer for PCR
 CC amplification of a full length OMP106 DNA. The DNA fragment can also be
 CC used as a probe for screening M. catarrhalis genomic libraries for OMP106
 CC polypeptide coding sequences. The encoded peptide fragment can
 CC specifically bind to an antibody that binds the OMP106 polypeptide. The
 CC antibody is a cytotoxic antibody which mediates complement killing of
 CC M. catarrhalis. The OMP106 polypeptide, and its peptide fragments can be
 CC used in vaccines and antigenic compositions. They can also be used for
 CC producing an immune response in an animal against M. catarrhalis.

CC Sequence 24 AA;

Query Match 57.9%; Score 121; DB 18; Length 24;
 Best Local Similarity 100.0%; Pred. No. 8.8e-09;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 EADGKGGANARCKSAIGDIAQ 28
 DB 1 aadgkkgganargdksiaigdiag 24

RESULT 6

ID AAE00702 standard; peptide: 24 AA.

AC AAE00702;

DT 02-JUL-2001 (first entry)

DE Internal fragment of N-terminal outer membrane protein-106 (OMP106).

KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;

KM bacterial infection; immunogen; cytotoxic; antituberc;

OS Moraxella catarrhalis.

XX 26-MAR-1996; 96US-0621944.
PR 01-MAY-1995; 95US-0431718.
PR 07-JUN-1995; 95US-0478370.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;
XR WPI; 1996-506162/50.
DR N-PsDB; AAT38740.
XX
PT Moraxella outer membrane protein - useful as immunogen in protective
PS vaccine and for diagnosis
XX
XX Claim 14; Fig 6; 109pp; English.

CC An approx. 200 kDa outer membrane protein (AAM04505) can be
CC isolated from *Moraxella catarrhalis* otitis media strain 4223
CC by electroporation, or expressed from a gene (see also AAT38740)
CC obdL from a strain 4223 genomic library. Natural or recombinant
CC outer membrane protein is useful as an immunogen to protect
CC against infection by *Moraxella*, esp. *M. catarrhalis*. It can
CC also be used to detect antibodies, esp. for differential diagnosis
CC between bacteria that cause similar symptoms, and also useful as
CC a carrier for other antigens and used to raise antitumour
CC antibodies for conjugation to therapeutic agents.

Sequence 1992 AA;
SQ

```

Query Match: 43.1%; Score 90; DB 17; Length 1992;
Best Local Similarity 52.3%; Pred. No. 0.012;
Matches: 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1

```

| RESULT | 8 |
|---|---|
| AA69133 | standard; Protein; 1992 AA. |
| AA69133 | |
| AA69133 | |
| 24-APR-2001 | (first entry) |
| M. catarrhalis strain 4223 | lambdaEMBL3 clone 200kda protein SEQ ID NO:33 |
| Moraxella catarrhalis strain 4223 | major outer membrane protein; |
| 200kda outer membrane protein; | antibacterial; immunogenic; infection; |
| ocitis media; detection. | |
| Moraxella catarrhalis. | |
| MO200107619-A1. | |
| 01-FEB-2001. | |
| 26-JUL-2000 | 2000MO-CA00870. |
| 27-JUL-1999 | 990US-0361619. |
| (CONN-) CONNAUGHT LAB LTD. | |
| Loomore SM, Sasaki K, Yang Y, Klein MH; | |
| WPI: 2001-159722/16. | |
| N-FSDB; AAF59100, AAF59101. | |
| New nucleic acid encoding Moraxella catarrhalis outer membrane protein; | |
| useful in protective vaccines and for diagnosis | |

PS Example 3; Fig 2A-W; 247pp; English.

CC The present invention describes an isolated and purified nucleic acid (I)

CC that encodes a 200 kDa outer membrane protein of *Moraxella catarrhalis*.

CC The 200 kDa outer membrane protein (II) has antibacterial activity and

CC can be used in vaccines. (II), and its truncated versions, are used as

CC immunogenic compositions and vaccines to protect against *M. catarrhalis*

CC infections, particularly otitis media in humans. (II) is also used as

CC antigen in immunoassays for detecting specific antibodies (Ab), and to

CC generate Ab. (I) are used for recombinant production of (II) and its

CC fragments are used as probes for identifying/cloning 200 kDa protein

CC genes from other strains, and for diagnostic detection of *M. catarrhalis*.

CC (I) makes possible production of large amount of recombinant immunogens.

CC Expression of truncated versions of (II) reduces toxicity of the protein

CC towards the *Escherichia coli* host. The present sequence represents the

CC *M. catarrhalis* strain 4223 lambdaDPMB3 clone 200kDa protein, which is

CC used in the exemplification of the present invention.

SO Sequence 1992 AA;

Query Match 43.1%; Score 90; DB 22; Length 1992;

Best Local Similarity 52.3%; Pred. No. 0.012;

Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

OY 4 SEADGKGGANAR---GDKSIAIGDIAQLGSQSIAIGDNKIV 43

DB 1718 akadgaavaiqrqtqngsialqdnagatgdsialgtgnv 1761

RESULT 9

AAB69137

ID AAB69137 standard; Protein: 1992 AA.

XX AAB69137;

DT 24-APR-2001 (first entry)

DE M. catarrhalis M56 200kDa protein in PK3348 SEQ ID NO:13.

XX Moraxella catarrhalis strain Q8; major outer membrane protein;

KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

XX otitis media; detection.

OS Moraxella catarrhalis.

XX WO200107619-A1.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-CA00870.

PR 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNAUGHT LAB LTD.

PI Loosmore SM, Sasaki K, Yang Y, Klein MH;

DR WPI; 2001-159722/16.

XX N-PSDB; AAF59106.

PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

XX useful in protective vaccines and for diagnosis

PS Claim 1; Fig 8A-V; 247pp; English.

CC The present invention describes an isolated and purified nucleic acid (I)

CC that encodes a 200 kDa outer membrane protein of *Moraxella catarrhalis*.

CC The 200 kDa outer membrane protein (II) has antibacterial activity and

CC can be used in vaccines. (II), and its truncated versions, are used as

CC immunogenic compositions and vaccines to protect against *M. catarrhalis*

CC infections, particularly otitis media in humans. (II) is also used as

CC antigen in immunoassays for detecting specific antibodies (Ab), and to

CC generate Ab. (I) are used for recombinant production of (II) and its

CC fragments are used as probes for identifying/cloning 200 kDa protein

CC genes from other strains, and for diagnostic detection of *M. catarrhalis*.

CC (I) makes possible production of large amount of recombinant immunogens.

CC Expression of truncated versions of (II) reduces toxicity of the protein

CC towards the *Escherichia coli* host. The present sequence represents the

CC *M. catarrhalis* M56/200kDa protein in PK3348, which is given in the

CC exemplification of the present invention.

SO Sequence 1992 AA;

Query Match 43.1%; Score 90; DB 22; Length 1992;

Best Local Similarity 52.3%; Pred. No. 0.012;

Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

OY 4 SEADGKGGANAR---GDKSIAIGDIAQLGSQSIAIGDNKIV 43

DB 1718 akadgaavaiqrqtqngsialqdnagatgdsialgtgnv 1761

RESULT 10

AAB69134

ID AAB69134 standard; Protein: 2047 AA.

XX AAB69134;

DT 24-APR-2001 (first entry)

DE M. catarrhalis strain 4223 genomic 200kDa protein SEQ ID NO:7.

XX Moraxella catarrhalis strain 4223; major outer membrane protein;

KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

XX otitis media; detection.

OS Moraxella catarrhalis.

XX WO200107619-A1.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-CA00870.

PR 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNAUGHT LAB LTD.

PI Loosmore SM, Sasaki K, Yang Y, Klein MH;

DR WPI; 2001-159722/16.

XX N-PSDB; AAF59102, AAF59103.

PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

XX useful in protective vaccines and for diagnosis

PS Claim 1; Fig 3A-W; 247pp; English.

CC The present invention describes an isolated and purified nucleic acid (I)

CC that encodes a 200 kDa outer membrane protein of *Moraxella catarrhalis*.

CC The 200 kDa outer membrane protein (II) has antibacterial activity and

CC can be used in vaccines. (II), and its truncated versions, are used as

CC immunogenic compositions and vaccines to protect against *M. catarrhalis*

CC infections, particularly otitis media in humans. (II) is also used as

CC antigen in immunoassays for detecting specific antibodies (Ab), and to

CC generate Ab. (I) are used for recombinant production of (II) and its

CC fragments are used as probes for identifying/cloning 200 kDa protein

CC genes from other strains, and for diagnostic detection of *M. catarrhalis*.

CC (I) makes possible production of large amount of recombinant immunogens.

CC Expression of truncated versions of (II) reduces toxicity of the protein

CC towards the *Escherichia coli* host. The present sequence represents the

CC *M. catarrhalis* strain 4223 genomic 200kDa protein, which is given in the

CC exemplification of the present invention.

SQ Sequence 2047 AA;

Query Match 43.1%; Score 90; DB 22; Length 2047;

Best Local Similarity 52.3%; Pred. No. 0.013; Mismatches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

OY 4 SEADGKGGANAR---GDKSIAIGDIAQALGSGSIAIGDNKIV 43
 :||| | | :||| | | | | | | :|
 Db 1773 akadgaavaigrtgagngslagdngatgdsalgtgnv 1816

RESULT 11

ID AAB69135 standard; Protein; 2053 AA.

AC AAB69135;

DT 24-APR-2001 (first entry)

DE M. catarrhalis strain Q8 200kDa protein SEQ ID NO:9.

KM Moraxella catarrhalis strain Q8; major outer membrane protein;
 KM 200kDa outer membrane protein; antibacterial; immunogenic; infection;
 KM otitis media; detection.

XX Moraxella catarrhalis.

XX MO200107619-A1.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000MO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

XX N-PSDB; AAF59104.

PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
 useful in protective vaccines and for diagnosis
 PS Claim 1; Fig 4A-V; 247pp; English.

CC The present invention describes an isolated and purified nucleic acid (I)
 CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
 CC The 200 kDa outer membrane protein (II) has antibacterial activity and
 CC can be used in vaccines. (II), and its truncated versions, are used as
 CC immunogenic compositions and vaccines to protect against M. catarrhalis
 CC infections, particularly otitis media in humans. (II) is also used as
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to
 CC generate Ab. (I) are used for recombinant production of (II) and its
 CC fragments are used as probes for identifying/cloning 200 kDa protein
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.
 CC (I) makes possible production of large amount of recombinant immunogens.
 CC Expression of truncated versions of (II) reduces toxicity of the protein
 CC towards the Escherichia coli host. The present sequence represents the
 CC M. catarrhalis strain Q8 200kDa protein, which is given in the
 CC exemplification of the present invention.

SQ Sequence 2053 AA;

Query Match 43.1%; Score 90; DB 22; Length 2053;

Best Local Similarity 52.3%; Pred. No. 0.013; Mismatches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

OY 4 SEADGKGGANAR---GDKSIAIGDIAQALGSGSIAIGDNKIV 43
 :||| | | :||| | | | | | | :|

Db 1779 akadgaavaigrtgagngslagdngatgdsalgtgnv 1822

RESULT 12

ID AAB69136 standard; Protein; 2314 AA.

AC AAB69136;

DT 24-APR-2001 (first entry)

DE M. catarrhalis les1 200kDa protein SEQ ID NO:11.

KM Moraxella catarrhalis strain Q8; major outer membrane protein;
 KM 200kDa outer membrane protein; antibacterial; immunogenic; infection;
 KM otitis media; detection.

XX Moraxella catarrhalis.

XX MO200107619-A1.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000MO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

XX N-PSDB; AAF59105.

PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
 useful in protective vaccines and for diagnosis
 PS Claim 1; Fig 5A-Y; 247pp; English.

CC The present invention describes an isolated and purified nucleic acid (I)
 CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
 CC The 200 kDa outer membrane protein (II) has antibacterial activity and
 CC can be used in vaccines. (II), and its truncated versions, are used as
 CC immunogenic compositions and vaccines to protect against M. catarrhalis
 CC infections, particularly otitis media in humans. (II) is also used as
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to
 CC generate Ab. (I) are used for recombinant production of (II) and its
 CC fragments are used as probes for identifying/cloning 200 kDa protein
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.
 CC (I) makes possible production of large amount of recombinant immunogens.
 CC Expression of truncated versions of (II) reduces toxicity of the protein
 CC towards the Escherichia coli host. The present sequence represents the
 CC M. catarrhalis les1 200kDa protein, which is given in the exemplification
 CC of the present invention.

SQ Sequence 2314 AA;

Query Match 43.1%; Score 90; DB 22; Length 2314;

Best Local Similarity 52.3%; Pred. No. 0.014; Mismatches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

OY 4 SEADGKGGANAR---GDKSIAIGDIAQALGSGSIAIGDNKIV 43
 :||| | | :||| | | | | | | :|
 Db 2040 akadgaavaigrtgagngslagdngatgdsalgtgnv 2083

RESULT 13

ID ABB52677 standard; Protein; 1778 AA.

XX ABB52677;
 AC ABB52677;

DE 11-FEB-2002 (first entry)
 XX Escherichia coli polypeptide SEQ ID NO 749.
 DE Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
 XX immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 KW systemic infection; non-diarrhoeal infection; septicemia;
 KW pyelonephritis; antibiotic resistance.
 XX Escherichia coli.
 OS WO200166572-A2.
 PN 13-SEP-2001.
 XX 12-MAR-2001; 2001WO-EP03445.
 PF 10-MAR-2000; 2000FR-0003145.
 PR 02-FEB-2001; 2001FR-0001449.
 XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
 DR WPI: 2001-550253/61.
 XX A library of DNA fragments of Escherichia coli strains for the
 PT phylogenetic determination of a given strain comprises polynucleotides of
 PT nature B2/D+ A -
 XX Example 6; Fig 6; 646pp: English.
 XX The invention relates to a library of DNA fragments of Escherichia coli
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
 CC and encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature
 CC B2/D+A-. The polynucleotides have potential antiinflammatory,
 CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli
 CC infections. The polypeptides are useful for determining the phylogenic
 CC group of a given E. coli strain. These polypeptides can detect and treat
 CC an undesired development of E. coli, particularly an extra-intestinal
 CC infection that include systemic and non-diarrhoeal infections such as
 CC septicemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more
 CC frequent use of broad spectrum antibiotics.
 CC Sequence 1778 AA;
 SQ

Query Match 32.5%; Score 68; DB 22; Length 1778;
 Best Local Similarity 53.8%; Pred. No. 6.8;
 Matches 14; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
 OY 15 ARGDSIAIGDIAQALGSOSIAIGD 40
 DB 170 adgkctialgnkayelmalsalgn 195

RESULT 14
 AAM56319
 ID AAM56319 standard; Protein; 2042 AA.
 AC AAM56319;
 XX 19-AUG-1998 (first entry)
 DE Haemophilus paragallinarum antigenic protein#1.
 XX Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;
 KW vaccine; chicken infectious coryza; CIC; fowl.
 XX Haemophilus paragallinarum.
 OS (GETH) GENENTECH INC.

FH Key Location/Qualifiers
 FT Peptide 1..70 /label= signal
 FT Protein 71..2042 /note= "antigenic protein"
 XX WO9812331-A1.
 XX 26-MAR-1998.
 PD 12-SEP-1997; 97WO-JF03222.
 XX 19-SEP-1996; 96JP-0271408.
 PR (KAGA) ZH KAGAKU & KESSEI ROHO KENKYUSHO.
 PA (KAGA) CHERO-SERO-THERAPEUTIC RES INST.
 PI Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;
 DR WFI: 1998-230318/20.
 DR N-FSDB; AAV22834.
 XX Antigenic polypeptide from Haemophilus paragallinarum induces HI
 PT antibody production - and is useful for diagnosis of and preparation
 PT of vaccines for chicken infectious coryza
 XX Claim 1; Page 51-69; 108pp; Japanese.
 XX The present sequence represents an antigenic protein derived from
 CC Haemophilus paragallinarum strain A-221. The antigenic protein
 CC stimulates the production of HI antibodies in fowl. The protein
 CC and DNA coding for it can be used in the preparation of vaccines
 CC for the prevention of chicken infectious coryza (CIC). The protein
 CC and its antibodies can be used in the diagnosis and treatment of CIC.
 CC Sequence 2042 AA;
 SQ

Query Match 32.1%; Score 67; DB 19; Length 2042;
 Best Local Similarity 45.2%; Pred. No. 11;
 Matches 14; Conservative 8; Mismatches 9; Indels 0; Gaps 0;
 OY 9 GKGANARGDSIAIGDIAQALGSOSIAIGD 39
 DB 322 gkigaialgrtiaaemstavgslafalaad 352

RESULT 15
 AAY13486
 ID AAY13486 standard; Protein; 261 AA.
 AC AAY13486;
 XX 26-JUL-1999 (first entry)
 DE Mouse clone 65 protein.
 XX Human; murine; clone 65; clone 320; Wnt; arteriosclerosis; malignancy;
 KW melanoma; cancer; breast; ovary; colon; tumor; cardiac; renal;
 KW inflammation; angiogenesis; immunological disorder.
 OS Mus sp.
 XX WO9921999-A2.
 PN 06-MAY-1999.
 PD 29-OCT-1998; 98WO-US22992.
 XX 04-FEB-1998; 98US-0073612.
 PR 29-OCT-1997; 97US-0063704.
 XX (GETH) GENENTECH INC.

XX Levine AJ, Pennica D;
 XX WPI: 1999-326705/27.
 DR N-PSDB; AAX55608, AAX55609.
 XX

PT Human and murine genes induced by Wnt and derived polypeptides

PS Claim 6; Fig 1A-B; 114pp; English.

XX
 CC The invention relates to human and murine cDNA clones 65 and 320,
 CC encoding polypeptides that are induced by Wnt. Clone 65 and 320
 CC polypeptides, and their antagonists, are used to treat disorders
 CC associated with these polypeptides, e.g. arteriosclerosis, malignancies
 CC (particularly melanoma or cancer of the breast, ovary and colon), but
 CC also benign tumors, cardiac, renal, catabolic or many other types of
 CC disease. Inflammation, angiogenesis and immunological disorders, more
 CC generally to induce death of Wnt-induced cells. Antibodies against the
 CC polypeptides are used as therapeutic antagonists and as diagnostic
 CC immunoassay or affinity-purification reagents. Clone 65 and 320 nucleic
 CC acids are used as hybridization probes or primers (to detect related
 CC sequences); for chromosome and gene mapping; to generate antisense
 CC sequences; for recombinant polypeptide production and to generate
 CC transgenic or 'knockout' animals (for development of, and screening for,
 CC drugs). The polypeptides are used to raise (or purify) specific
 CC antibodies; as immunoassay reagents and in drug screens.
 XX

SQ Sequence 261 AA;

Query Match

30.1%; Score 63; DB 20; Length 261;

Best Local Similarity 38.9%; Pred. No. 3.3;

Matches 14; Conservative 6; Mismatches 14; Indels 2; Gaps 1;

OY 2 GISEADGGKGGANARGPDKSIAIGDIAQALGSOSIAI 37

DB 37 gvsgrgraggaeqrqvcvlygd--gavgkcsivv 70

Search completed: July 30, 2002, 15:34:11
 Job time: 406 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:30:55 ; Search time 33.83 Seconds
(without alignments)
31.046 Million cell updates/sec

Title: US-09-813-214A-1

Perfect score: 209
Sequence: 1 IGISEADGKGANARGDKS.....GDIQAALSGSIAIDNKIV 43

Scoring table: BLOSOM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/backfill1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 209 | 100.0 | 43 | 4 | US-08-968-685A-1 |
| 2 | 209 | 100.0 | 2123 | 4 | US-08-968-685A-10 |
| 3 | 190 | 90.9 | 40 | 4 | US-08-968-685A-12 |
| 4 | 121 | 57.9 | 24 | 4 | US-08-968-685A-5 |
| 5 | 121 | 57.9 | 24 | 4 | US-08-968-685A-13 |
| 6 | 121 | 57.9 | 24 | 4 | US-08-968-685A-19 |
| 7 | 90 | 43.1 | 2314 | 4 | US-09-268-347-49 |
| 8 | 89 | 42.6 | 2048 | 4 | US-09-268-347-48 |
| 9 | 60 | 28.7 | 278 | 1 | US-08-188-582-9 |
| 10 | 60 | 28.7 | 278 | 1 | US-08-646-715-9 |
| 11 | 55 | 26.3 | 379 | 4 | US-09-180-827-12 |
| 12 | 55 | 26.3 | 889 | 4 | US-09-336-447A-15 |
| 13 | 53.5 | 25.6 | 335 | 2 | US-08-405-175A-6 |
| 14 | 52.5 | 25.1 | 1160 | 3 | US-08-808-559A-24 |
| 15 | 52 | 24.9 | 513 | 1 | US-08-173-508-6 |
| 16 | 52 | 24.9 | 513 | 2 | US-08-265-310-6 |
| 17 | 52 | 24.9 | 513 | 3 | US-08-951-742-6 |
| 18 | 51 | 24.4 | 580 | 2 | US-08-591-079-2 |
| 19 | 50.5 | 24.2 | 126 | 3 | US-07-765-830A-1 |
| 20 | 50.5 | 24.2 | 126 | 3 | US-07-765-830A-6 |
| 21 | 50.5 | 24.2 | 126 | 3 | US-07-765-830A-7 |
| 22 | 50.5 | 24.2 | 768 | 2 | US-08-222-617A-5 |
| 23 | 50.5 | 24.2 | 3666 | 2 | US-08-222-617A-12 |
| 24 | 50.5 | 24.2 | 3727 | 2 | US-08-222-617A-27 |
| 25 | 50.5 | 24.2 | 3778 | 2 | US-08-222-617A-2 |
| 26 | 50 | 23.9 | 249 | 2 | US-08-797-689-18 |
| 27 | 50 | 23.9 | 1002 | 4 | US-09-268-347-24 |

| | | | | | | |
|----|------|------|------|---|--------------------|-------------------|
| 28 | 50 | 23.9 | 1004 | 4 | US-09-268-347-30 | Sequence 30, Appl |
| 29 | 49.5 | 23.7 | 430 | 1 | US-08-535-237-2 | Sequence 2, Appl |
| 30 | 49.5 | 23.7 | 892 | 4 | US-09-336-447A-5 | Sequence 5, Appl |
| 31 | 49 | 23.4 | 126 | 2 | US-07-757-606B-1 | Sequence 1, Appl |
| 32 | 49 | 23.4 | 126 | 2 | US-07-757-606B-6 | Sequence 6, Appl |
| 33 | 49 | 23.4 | 126 | 3 | US-07-728-220C-20 | Sequence 20, Appl |
| 34 | 49 | 23.4 | 126 | 3 | US-08-838-151A-20 | Sequence 20, Appl |
| 35 | 49 | 23.4 | 624 | 4 | US-09-336-447A-9 | Sequence 22, Appl |
| 36 | 48.5 | 23.2 | 248 | 2 | US-08-887-352B-22 | Sequence 22, Appl |
| 37 | 48.5 | 23.2 | 248 | 2 | US-08-887-352B-23 | Sequence 22, Appl |
| 38 | 48.5 | 23.2 | 248 | 4 | US-09-109-207C-22 | Sequence 22, Appl |
| 39 | 48.5 | 23.2 | 248 | 4 | US-09-109-207C-23 | Sequence 22, Appl |
| 40 | 48.5 | 23.2 | 248 | 4 | US-09-296-005-22 | Sequence 22, Appl |
| 41 | 48.5 | 23.2 | 248 | 4 | US-09-296-005-23 | Sequence 22, Appl |
| 42 | 48.5 | 23.2 | 941 | 4 | US-09-336-447A-9 | Sequence 9, Appl |
| 43 | 48 | 23.0 | 240 | 1 | US-08-488-113B-148 | Sequence 148, App |
| 44 | 48 | 23.0 | 240 | 1 | US-08-477-484B-148 | Sequence 148, App |
| 45 | 48 | 23.0 | 240 | 2 | US-08-646-360-148 | Sequence 148, App |

ALIGNMENTS

RESULT 1:
; Sequence 1, Application US/08968685A
; Patent No. 6214981
; GENERAL INFORMATION:
; APPLICANT: PLOSILA, KENNETH
; APPLICANT: PLOSILA, LAURA
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS: EDMONDS LLP
; ADDRESS: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968,685A
; FILING DATE: No. 6214981ember 12, 1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7969-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-968-685A-1

Query Match 100.0%; Score 209; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 3.6e-20;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 IGISEADGKGANARGDKSIAIDNKIV 43
|||||

Db 1 IGISEADGKGANARGDKSIAIGDIAQALGSQSIAIGDNKIV 43

RESULT 2

US-08-968-685A-10
Sequence 10, Application US/08968685A
Patent No. 6214981

GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH
APPLICANT: PLOSILA, LAURA
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,685A
FILING DATE: No. 6214981ember 12, 1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2123 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-968-685A-10

Query Match 100.0%; Score 209; DB 4; Length 2123;
Best Local Similarity 100.0%; Pred. No. 3.3e-18;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IGISEADGKGANARGDKSIAIGDIAQALGSQSIAIGDNKIV 43
Db 69 IGISEADGKGANARGDKSIAIGDIAQALGSQSIAIGDNKIV 111

RESULT 3

US-08-968-685A-12
Sequence 12, Application US/08968685A
Patent No. 6214981

GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH
APPLICANT: PLOSILA, LAURA
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,685A
FILING DATE: No. 6214981ember 12, 1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-968-685A-12

Query Match 90.9%; Score 190; DB 4; Length 40;
Best Local Similarity 100.0%; Pred. No. 8.5e-18;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IGISEADGKGANARGDKSIAIGDIAQALGSQSIAIGD 39
Db 2 IGISEADGKGANARGDKSIAIGDIAQALGSQSIAIGD 40

RESULT 4

US-08-968-685A-5
Sequence 5, Application US/08968685A
Patent No. 6214981

GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH
APPLICANT: PLOSILA, LAURA
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,685A
FILING DATE: No. 6214981ember 12, 1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-968-685A-5

Query Match 57.9%; Score 121; DB 4; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.7e-09;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EADGKGANARGDKSIAIGDIAQ 28
DB 1 EADGKGANARGDKSIAIGDIAQ 24

RESULT 5
US-08-968-685A-13
Sequence 13, Application US/08968685A
Patent No. 6214981
GENERAL INFORMATION:
APPLICANT: TUCKER, KENNETH
APPLICANT: PLOSTILA, LAURA
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,685A
FILING DATE: No. 6214981ember 12, 1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-968-685A-13

Query Match 57.9%; Score 121; DB 4; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.7e-09;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EADGKGANARGDKSIAIGDIAQ 28
DB 1 EADGKGANARGDKSIAIGDIAQ 24

RESULT 6
US-08-968-685A-19
Sequence 19, Application US/08968685A
Patent No. 6214981
GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH
APPLICANT: PLOSTILA, LAURA
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,685A
FILING DATE: No. 6214981ember 12, 1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-968-685A-19

Query Match 57.9%; Score 121; DB 4; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.7e-09;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EADGKGANARGDKSIAIGDIAQ 28
DB 1 EADGKGANARGDKSIAIGDIAQ 24

RESULT 7
US-09-268-347-49
Sequence 49, Application US/09268347
Patent No. 6335182
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
FILE REFERENCE: 1038-860
CURRENT APPLICATION NUMBER: US/09/268,347
CURRENT FILING DATE: 1999-03-16
NUMBER OF SEQ ID NOS: 54
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 49
LENGTH: 2314
TYPE: PRP
ORGANISM: Haemophilus influenzae
US-09-268-347-49

Query Match 43.1%; Score 90; DB 4; Length 2314;
Best Local Similarity 52.3%; Pred. No. 0.0047;
Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

QY 4 SEADGKGANAR---GDKSIAIGDIAIGSOSIAIGDKIV 43
:||||| | :||||||| ||| | ||||| :|

DB 2040 AKADGEAAVAIGRGTQAGNQSIAIGDMAQATGDSIAIGTGNV 2083

RESULT 8
US-09-268-347-48
Sequence 48, Application US/09268347
Patent No. 6335182
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
FILE REFERENCE: 1038-860
CURRENT APPLICATION NUMBER: US/09/268,347
CURRENT FILING DATE: 1999-03-16
NUMBER OF SEQ ID NOS: 54
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 48
LENGTH: 2048
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-268-347-48

Query Match 42.6%; Score 89; DB 4; Length 2048;
Best Local Similarity 52.3%; Pred. No. 0.0055;
Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

OY 4 SEADGKGKANAR---GDKSIAIGDIAQALGQSIAIGDKIV 43
DB 1774 AKADGEAAVAIGRGTQAGNQSIAIGDMAQATGDSIAIGTGNV 1817

RESULT 9
US-08-188-582-9
Sequence 9, Application US/08188582
Patent No. 5534410
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
NUMBER OF SEQUENCES: 36
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELECOMMUNICATION INFORMATION:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:

LENGTH: 278 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-188-582-9

Query Match 28.7%; Score 60; DB 1; Length 278;
Best Local Similarity 31.6%; Pred. No. 2.6;
Matches 12; Conservative 8; Mismatches 18; Indels 0; Gaps 0;

OY 1 IGISEADGKGKANARGDKSIATIGDIAQALGQSIAIG 38
DB 198 VGSSGGSGGGGGGQEVKESSTGAGDLMKEVDSDAANG 235

RESULT 10
US-08-646-715-9
Sequence 9, Application US/08646715
Patent No. 5637686
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
NUMBER OF SEQUENCES: 36
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELECOMMUNICATION INFORMATION:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-646-715-9

Query Match 28.7%; Score 60; DB 1; Length 278;
Best Local Similarity 31.6%; Pred. No. 2.6;
Matches 12; Conservative 8; Mismatches 18; Indels 0; Gaps 0;

OY 1 IGISEADGKGKANARGDKSIATIGDIAQALGQSIAIG 38

Db 198 VGSSGGSGGGGDEYKSESTGAGGDLMEVDSDAAAVG 235

RESULT 11

US-09-180-827-12

Sequence 12, Application US/09180827
Patent No. 6355464
GENERAL INFORMATION:
APPLICANT: Healy, Judith M.
APPLICANT: Bodorova, Jana
APPLICANT: Lam, Kelvin T.
APPLICANT: Lesson, Andrea J.
TITLE OF INVENTION: M. Tuberculosis RNA Polymerase Alpha
FILE REFERENCE: 0342/1C382-US2
CURRENT APPLICATION NUMBER: US/09/180,827
CURRENT FILING DATE: 1999-01-26
PRIOR APPLICATION NUMBER: PCT/US97/22216
PRIOR FILING DATE: 1997-11-03
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 12
LENGTH: 379
TYPE: PRF
ORGANISM: C. trachomatis
US-09-180-827-12

Query Match

26.3% Score 55; DB 4; Length 379;
Best Local Similarity 27.0%; Pred. No. 16;
Matches 10; Conservative 9; Mismatches 18; Indels 0; Gaps 0;

QY 3 ISEADGGKGANARGDKSIAIGDIAQALGSGSIAIGD 39

Db 106 LDCEGRCRCKLRATISVDASDLAAAGGKEVTIGD 142

RESULT 12

US-09-336-447A-15

Sequence 15, Application US/09336447A
Patent No. 6310190
GENERAL INFORMATION:
APPLICANT: HANSEN, ERIC J.
APPLICANT: AEBL, CHRISTOPH
APPLICANT: COPE, LESLIE D.
APPLICANT: MACIVER, ISOBEL
APPLICANT: FISKE, MICHAEL J.
APPLICANT: FREDENBURG, ROSS A.
TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
FILE REFERENCE: AMCY:024
CURRENT APPLICATION NUMBER: US/09/336,447A
CURRENT FILING DATE: 1999-06-21
NUMBER OF SEQ ID NOS: 98
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 15
LENGTH: 889
TYPE: PRF
ORGANISM: Moraxella catarrhalis
US-09-336-447A-15

Query Match

26.3% Score 55; DB 4; Length 889;
Best Local Similarity 34.1%; Pred. No. 43;
Matches 14; Conservative 5; Mismatches 16; Indels 6; Gaps 1;

QY 9 GKGG-----ANARGDKSIAIGDIAQALGSGSIAIGDKNIV 43

Db 320 GKGSFAGIDNKNANADNAVALGNKNTIEGNSVALGSNNIV 360

RESULT 13

US-08-405-175A-6

Sequence 6, Application US/08405175A
Patent No. 5885772

GENERAL INFORMATION:
APPLICANT: Aderem, Alan A.
APPLICANT: Chen, Jianmin
APPLICANT: Chang, Sandy
TITLE OF INVENTION: METHOD FOR THE DETECTION OF ANENCEPHALY
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,175A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-121A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 335 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: predicted primary structure of bovine MARCKS
HYPOTHEICAL: NO
US-08-405-175A-6

Query Match

25.6% Score 53.5; DB 2; Length 335;
Best Local Similarity 42.9%; Pred. No. 21;
Matches 15; Conservative 3; Mismatches 16; Indels 1; Gaps 1;

QY 5 EADGGKGANARGDKSIAIGDIAQALGSGSIAIGD 39

Db 182 EAEGA-AGASAGSGKDEASGGAAGAAAGAGAPGE 215

RESULT 14

US-08-808-599A-24

Sequence 24, Application US/08808599A
Patent No. 6111089
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin, Trophinin-Assisting
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/808.599A
FILING DATE: 28-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/317,522
FILING DATE: 04-OCT-1994
APPLICATION DATA:
APPLICATION NUMBER: US 08/439,818
FILING DATE: 12-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-1A 2256
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1160 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-808-599A-24

```

```

Query Match      25.1%; Score 52.5; DB 3; Length 1160;
Best Local Similarity 38.1%; Pred. No. 1.2e+02;
Matches 16; Conservative 3; Mismatches 14; Indels 9; Gaps 2;

OY 2 GISEADGKGANARGDKSIAIGDI---AQLGSGSIAIGD 39
Db 171 GISNPSGGGGRN-----SIFGVSVPNTSANSAPSISFGD 207

```

```

RESULT 15
US-08-173-508-6
Sequence 6, Application US/08173508
Patent No. 5616485
GENERAL INFORMATION:
APPLICANT: Bartfeld, Daniel
APPLICANT: Butler, Michael J.
APPLICANT: Hadary, Dany
APPLICANT: Jenish, David
APPLICANT: Krieger, Timothy
TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/173,508
FILING DATE: 23-DEC-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 18740/125/CACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136

```

```

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 513 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-173-508-6

```

```

Query Match      24.9%; Score 52; DB 1; Length 513;
Best Local Similarity 32.4%; Pred. No. 54;
Matches 11; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

OY 2 GISEADGKGANARGDKSIAIGDIAQALGSGSI 35
Db 24 GSDPDGDDGGRSSAGPSAAPSVPALASQTL 57

```

Search completed: July 30, 2002, 15:34:51
 Job time: 236 sec

Wed Jul 31 07:54:03 2002

us-09-813-214a-1.ral

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 30, 2002, 15:31:35 ; Search time 51.53 Seconds
(without alignments)
80.183 Million cell updates/sec

Title: US-09-813-214A-1
Perfect score: 209
Sequence: 1 IGISEADGKGCANMARGDS.....GDLAQLGSGSIAIGDNKIV 43

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR71:*

1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 70.5 | 33.7 | 1588 | 2 A86036 | probable adhesin Z |
| 2 | 70.5 | 33.7 | 1588 | 2 H91188 | probable adhesin E |
| 3 | 66 | 31.6 | 1004 | 2 C82672 | surface-exposed on |
| 4 | 65.5 | 31.3 | 434 | 2 S04534 | invasin precursor |
| 5 | 65.5 | 31.3 | 658 | 2 AH0110 | probable surface P |
| 6 | 64.5 | 30.9 | 1190 | 2 A82615 | surface protein XF |
| 7 | 64 | 30.6 | 1107 | 2 AC0976 | probable autotrans |
| 8 | 63.5 | 30.4 | 455 | 2 S04912 | Yopa protein - Yer |
| 9 | 63 | 30.1 | 2059 | 2 D82671 | surface protein XF |
| 10 | 62.5 | 29.9 | 436 | 2 T36478 | probable solute-bl |
| 11 | 61 | 29.2 | 394 | 2 B81663 | DNA-directed RNA P |
| 12 | 61 | 29.2 | 615 | 2 H70589 | hypothetical glyci |
| 13 | 61 | 29.2 | 615 | 2 F70868 | hypothetical glyci |
| 14 | 60 | 28.7 | 278 | 2 A49067 | transcription init |
| 15 | 60 | 28.7 | 327 | 2 T00797 | hypothetical prote |
| 16 | 59.5 | 28.5 | 104 | 2 T41381 | Chaperonins 10 kD |
| 17 | 59.5 | 28.5 | 365 | 2 AB3486 | cell surface prote |
| 18 | 59 | 28.2 | 401 | 2 C88571 | protein C05B5.3 [1 |
| 19 | 58.5 | 28.0 | 562 | 2 T49788 | related to merozoi |
| 20 | 58 | 27.8 | 293 | 2 AB3639 | serine proteinase |
| 21 | 58 | 27.8 | 1176 | 2 T18042 | ice nucleation pro |
| 22 | 57.5 | 27.5 | 676 | 1 EDBE22 | immediate-early pr |
| 23 | 57.5 | 27.5 | 1660 | 2 A70869 | hypothetical glyci |
| 24 | 57 | 27.3 | 364 | 2 AE0169 | probable exported |
| 25 | 57 | 27.3 | 422 | 2 S04911 | Yopa protein - Yer |
| 26 | 57 | 27.3 | 505 | 2 AC3486 | cell surface prote |
| 27 | 56.5 | 27.0 | 1737 | 2 AC9235 | unconventional myo |
| 28 | 56.5 | 27.0 | 245 | 2 H69471 | conserved hypothet |
| 29 | 56.5 | 27.0 | 2342 | 2 T18200 | fatty-acid synthas |

| | | | | | |
|----|------|------|------|----------|----------------------|
| 30 | 56 | 26.8 | 377 | 2 A71505 | DNA-directed RNA P |
| 31 | 56 | 26.8 | 552 | 2 T23755 | hypothetical prote |
| 32 | 56 | 26.8 | 584 | 2 G70804 | hypothetical glyci |
| 33 | 55.5 | 26.6 | 698 | 2 C82332 | translational elonga |
| 34 | 55 | 26.3 | 229 | 2 AB0220 | flagellar L-ring P |
| 35 | 55 | 26.3 | 447 | 2 E83465 | conserved hypothet |
| 36 | 54.5 | 26.1 | 297 | 2 E97387 | hypothetical prote |
| 37 | 54.5 | 26.1 | 37 | 2 AF2605 | oxidoreductase Act |
| 38 | 54.5 | 26.1 | 340 | 2 T20807 | hypothetical prote |
| 39 | 54.5 | 26.1 | 467 | 2 F96773 | geranylgeranyl red |
| 40 | 54.5 | 26.1 | 562 | 2 B70953 | hypothetical glyci |
| 41 | 54.5 | 26.1 | 644 | 2 T15652 | hypothetical prote |
| 42 | 54.5 | 26.1 | 728 | 2 AF3299 | malate synthase 9 |
| 43 | 54.5 | 26.1 | 741 | 2 G70917 | hypothetical glyci |
| 44 | 54.5 | 26.1 | 1538 | 2 H70846 | hypothetical glyci |
| 45 | 54 | 25.8 | 196 | 2 G85435 | Tiny-like protein |

ALIGNMENTS

RESULT 1
A86036.
probable adhesin Z5029 [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: A86036
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A86036
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1588 <STO>
A:Cross-references: GB:AE005174; NID:912518349; PIDN:AAG58749.1; GSPDB:GN00145; UNCP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genet:73;
A:Gene: Z5029

Query Match 33.7%; Score 70.5; DB 2; Length 1588;
Best Local Similarity 32.8%; Pred. No. 3.4;
Matches 19; Conservative 9; Mismatches 11; Indels 19; Gaps 2;

QY 4 SHADG-----GKG-----GANNRGCKSIAIGDLAQLGSGSIAIGDNKI 42
DB 261 NKAQGVDAIALGNGSQRGLNTIALGTASMATGDKSLAIGSNSSANGINSVALGADSI 338

RESULT 2
H91188
probable adhesin ECS4480 [similarity] - Escherichia coli (strain O157:H7, substrain R
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 02-Nov-2001
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuwara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9
A:Reference number: A95629; MUID:21156231; PMID:11258796
A:Accession: H91188
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1588 <HAV>
A:Cross-references: GB:BA000007; PIDN:BA837903.1; PID:913363955; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECS4480

Query Match 33.7%; Score 70.5; DB 2; Length 1588;

Best Local Similarity 32.8%; Pred. No. 3.4;
Matches 19; Conservative 9; Mismatches 11; Indels 19; Gaps 2;

QY 4 SEADGK-----GKG-----GANARGDKSIADIDIAQALGSSOIAIDNKI 42
Db 281 NKADGVNIALGNSOSRGINTIALGTASNATGDKSIATLSSNSANGINSVALGADSI 338

RESULT 3
C82672

surface-exposed outer membrane protein XF1516 [imported] - Xylella fastidiosa (strain 9a)

C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: C82672

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1004 <SIM>

A:Cross-references: GB:AE003961; GB:AE003849; NID:g9106543; PIDN:AAF84325.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; B

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H

as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laiz

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marzeca, M.V.; Martins, F

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Saretelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF1516

Query Match 31.6%; Score 66; DB 2; Length 1004;

Best Local Similarity 50.0%; Pred. No. 7.3;

Matches 15; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 9 GKGANARGDKSIADIDIAQALGSSOIAIG 38

Db 576 GSGVSAIKDKSTATGASNAQAVGSSVALG 605

RESULT 4

S04534

invasin precursor - Yersinia pseudotuberculosis plasmid pIBI

C:Species: Yersinia pseudotuberculosis

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999

C:Accession: S04534; S04910

R:Rosqvist, R.; Skurnik, M.; Wolf-Watz, H.

Nature 334, 522-525, 1988

A:Title: Increased virulence of Yersinia pseudotuberculosis by two independent mutations

A:Reference number: S04534; MUID:88302441

A:Accession: S04534

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-434 <ROS>

A:Cross-references: EMBL:X12758; EMBL:X13883; NID:g48659; PIDN:CAA32088.1; PID:g48640

R:Skurnik, M.; Wolf-Watz, H.

Mol. Microbiol. 3, 517-529, 1989

A:Title: Analysis of the yopA gene encoding the YopI virulence determinants of Yersinia

A:Reference number: S04910; MUID:89343638

A:Accession: S04910

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-434 <SKU>

A:Cross-references: EMBL:X12758; EMBL:X13883; NID:g48659; PIDN:CAA32088.1; PID:g48640

C:Genetics:

A:Gene: yopA

A:Genome: plasmid

Query Match 31.3%; Score 65.5; DB 2; Length 434;

Best Local Similarity 46.5%; Pred. No. 3.6;

Matches 20; Conservative 3; Mismatches 11; Indels 9; Gaps 2;

QY 9 GKGANARGDK--SIADIDIAO-----ALGSSOIAIDNKI 42

Db 94 GAGGIANARAKDPRYSIAIGTAPAAKPAVAVAGSSIAIGVNSV 136

RESULT 5

probable surface protein (partial) YP00902 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C:Accession: AH0110

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tittball, R.W.; Holden, M.T.G.; Prentice, M

deno-Tarrag, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barril

Nature 413, 525-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AH0110

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-658 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC89747.1; PID:g15978974; GSPDB:GN00175

C:Genetics:

A:Gene: YP00902

Query Match 31.3%; Score 65.5; DB 2; Length 658;

Best Local Similarity 39.0%; Pred. No. 5.5;

Matches 16; Conservative 6; Mismatches 18; Indels 1; Gaps 1;

QY 2 GISEADGGKGGANARGDKSIADIDIAQALGSSOIAIDNKI 42

Db 480 GANSATGGAGSV-ASGNNSTAFSGAKATANSALANCV 519

RESULT 6

S04534

surface protein XF1981 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: A82615

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: A82615

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1190 <SIM>

A:Cross-references: GB:AE004017; GB:AE003849; NID:g9107083; PIDN:AAF84783.1; GSPDB:GN

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer

as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marzeca, M.V.; Martins

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira
M.; Tshahko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1981

Query Match 30.9%; Score 64.5; DB 2; Length 1190;
Best Local Similarity 44.4%; Pred. No. 13;
Matches 16; Conservative 7; Mismatches 12; Indels 1; Gaps 1;

QY 4 SEADGKGAN-ARGDSIAIGDIAQALGSOSIAIG 38
Db 191 SKTDGNTGATVAGLRSIAIGTARSOSDPAISIG 226

RESULT 7
AC0976
probable autotransporter sabb [imported] - Salmonella enterica subsp. enterica serovar
C:Species: Salmonella enterica subsp. enterica serovar typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: AC0976
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Mout, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AC0976
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1107 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD03303.1; PID:916504923; GSPDB:GN00176
C:Genetics:
A:Gene: sabb

Query Match 30.6%; Score 64; DB 2; Length 1107;
Best Local Similarity 38.5%; Pred. No. 14;
Matches 15; Conservative 7; Mismatches 15; Indels 2; Gaps 1;

QY 4 SEADGKGANARGDSIAIGDIAQALGSOSIAIGDNKI 42
Db 879 TMTDG--ADANAGADSVAGISGSIAMAAENVALCTNSV 915

RESULT 8
S04912
yopA protein - Yersinia enterocolitica plasmid pY6471/76
C:Species: Yersinia enterocolitica
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S04912
R:Skurnik, M.; Wolf-Watz, H.
Mol. Microbiol. 3, 517-529, 1989
A:Title: Analysis of the yopA gene encoding the YopI virulence determinants of Yersinia
A:Reference number: S04910; MUID:89343638
A:Accession: S04912
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-455 <SKU>
A:Cross-references: EMBL:X13882; NID:948606; PIDN:CAA32086.1; PID:948607
C:Genetics:
A:Gene: plasmid pY6471/76

Query Match 30.4%; Score 63.5; DB 2; Length 455;
Best Local Similarity 44.2%; Pred. No. 6.4;
Matches 19; Conservative 6; Mismatches 9; Indels 9; Gaps 2;

QY 9 GKGG--ANARGDSIAIGDIAQALGSOSIAIGDNKI 42
Db 59 GAGGLNANAGHISIAIGTATPAKANAAYVAGSIAIGTNSV 101

RESULT 9
D82671
surface protein XF1529 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: D82671
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A8515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: D82671
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-2059 <SIM>
A:Cross-references: GB:AE003982; GB:AE003849; NID:9106554; PIDN:AAF84338.1; GSPDB:GN
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Relbach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
Submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurmae, E.E.; La
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palzeri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tshahko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1529

Query Match 30.1%; Score 63; DB 2; Length 2059;
Best Local Similarity 36.7%; Pred. No. 34;
Matches 18; Conservative 4; Mismatches 19; Indels 8; Gaps 1;

QY 2 GISEADGKG-----GANARGDSIAIGDIAQALGSOSIAIGDNKI 42
Db 1731 GNYGDGATGSKAIAAGVGTQASGEAAAYVGSAAAGSKRSTAGRAAI 1779

RESULT 10
T36478
probable solute-binding protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36478
R:Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M
submitted to the EMBL data library, September 1999
A:Reference number: Z21607
A:Accession: T36478
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-436 <SEED>
A:Cross-references: EMBL:AL110470; PIDN:CAB54175.1; GSPDB:GN00070; SCOEDB:SCF85.18
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCF85.18

Query Match 29.9%; Score 62.5; DB 2; Length 436;
Best Local Similarity 35.8%; Pred. No. 8.1;
Matches 19; Conservative 5; Mismatches 16; Indels 13; Gaps 2;

QY 2 GISEADGK-----KGA-----NARGDSIAIGDIAQALGSOSIAIGDNKI 41

```
Db      283 G1QADGCGFEVSSCGLGTYMFPFVDGCKGDPSNAAGNPAQYLSISSKATDEEK   335
||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT  11
B81663
DNA-directed RNA polymerase, alpha chain TC0794 [imported] - Chlamydia muridarum (strain
C.Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
C.Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Sep-2000
C.Accession: B81663
R.Read: T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; Hickey,
, C.; Dodson, K.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28; 1397-1406, 2000
A.Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A.Reference number: AB1500; MUID:20150255
A.Accession: B81663
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-394 <TE>
A.Cross-references: GB:AE002347; GB:AE002160; NID:g7190815; PIDN:AAF39597.1; PID:g7190815
A.Experimental source: strain Nigg (MOpn)
C.Genetics:
A.Gene: TC0794
C.Superfamily: DNA-directed RNA polymerase alpha chain
```

| Query Match | Similarity | Score | DB | Length |
|-------------|------------|--------------|----|----------------------------------|
| Best Local | 32.4% | Pred. No. | 11 | |
| Matches | 12 | Conservative | 8 | Mismatches 17; Indels 0; Gaps 0; |

| QY | 3 | ISEADGCKGGAARBDKSTAI | TDINQALGSGSIAIGD | 39 |
|----|-----|----------------------|------------------|-----|
| Db | 123 | LQPCGGRRSQKRTATISDAS | DLAAAGGQKRTITVD | 159 |

RESULT 12
 H70589
 hypothetical glycine-rich protein RV2853 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 21-Jan-2000
 C:Accession: H70589
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987
 A:Accession: H70589
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-615 <COL>
 A:Cross-references: GB:Z95207; GB:AL123456; NID:g3261745; PIDN:CAB08453.1; PID:e315182;
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV2853
 C:Superfamily: unassigned collagens

| | | | | |
|-----------------------|-------|--------------|------|---------------------------------|
| Query Match | 29.2% | Score 61 | DB 2 | Length 615 |
| Best Local Similarity | 41.5% | Pred. No. 17 | | |
| Matches | 17 | Conservative | 4 | Mismatches 18; Indels 2; Gaps 1 |

| | | | |
|----|-----|--|-----|
| QY | 2 | GISEADGGKGANARSKDAIG--DIAQALGSSQSIADGN | 40 |
| | | : : : : | |
| Db | 408 | GGAGGNGGTGGAGAPGAGGAGGAKMDINSLGDNATVYGCN | 448 |

RESULT 13
F70868
hypothetical glycine-rich protein Rv2487c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: F70868.
R:Coie, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature, 393, 537-544, 1998
A:Authors: Squares, R., Salston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
A:Reference number: A0500; MUID:98295987
A:Accession: F70868
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residue: 1-694 <COL>
A:Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAAL6064.1; PID:g279
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV2487c
C:Superfamily: elastin

| | | | | | |
|----|---|-------|----------------|------------|-------------|
| | Query Match | 29.2% | Score 61; | DB 2; | Length 694; |
| | Best Local Similarity | 38.3% | Pred. No. 19; | | |
| | Matches 18; Conservative | 3; | Mismatches 16; | Indels 10; | Gaps 2 |
| OY | 2 GISEADGK-----GGANARQDKSIATGDIAQL-----GSOSTAIG 38 | | | | |
| | | | | | |
| Db | 292 GGAGCGDGTAAAGGGGGAAGGDDGVAAAGDAFFGGSDGNGSDGYAAG 338 | | | | |

RESULT 14
A49067
transcription initiation factor IID chain p42 - fruit fly (*Drosophila melanogaster*)
N:Alternate names: TATA-binding protein (TBP)-associated factor TAFII40; TFIID complex
C:Species: *Drosophila melanogaster*
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 17-Nov-2000
C:Accession: A49067; S42221
R:Goodrich, J.A.; Hoey, T.; Thut, C.J.; Admon, A.; Tjian, R.
Cell 75, 519-530, 1993
A:Title: Drosophila TAF-II40 interacts with both a VP16 activation domain and the basal
A:Reference number: A49067; MUID:94037099
A:Accession: A49067
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-278 <GCG>
A:Cross-references: GB:L29540; NID:g463048; PID:g463049
A:Note: Parts of this sequence were confirmed by peptide sequencing
R:Kokubo, T.; Gong, D.W.; Wootton, J.C.; Horikoshi, M.; Koeder, R.G.; Nakatani, Y.
Nature 367, 484-487, 1994
A:Title: Molecular cloning of *Drosophila* TFIID subunits.
A:Reference number: S42220; MUID:94150630
A:Accession: S42221
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-278 <KKK>
A:Cross-references: EMBL:U06458; NID:g458679; PIDN:AAC47347.1; PID:g458680
C:Genetics:
A:Gene: FlyBase:Taf40
A:Cross-references: FlyBase:Fbgn0000617
;Keywords: transcription initiation

| | | | | | |
|---------|-----------------------|---|---------------|------------|-----------------------|
| | Query Match | 28.7% | Score 60; | DB 2; | Length 278; |
| | Best Local Similarity | 31.6%; | Pred. No. 10; | | |
| Matches | 12; | Conservative | 8; | Mismatches | 18; Indels 0; Gaps 0; |
| QY | 1 | IGISEADGGKGCANARGDKSTAGIDGIAOALQSOSTATG | 38 | | |
| | : | : : : : : : | : | | |
| Db | 198 | VGSSSGGSGGGGGGVNKSVESTGAGSLCKEIVSDAAAVG | 235 | | |

RESULT 15
T00797
hypothetical protein At2g32710 [imported] - Arabidopsis thaliana
N/Alternate names: hypothetical protein F24L7.15

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999, #text_change 16-Feb-2001
C:Accession: T00797; E84736
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, February 1998
A:Description: Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.
A:Reference number: Z14204
A:Accession: T00797
A:Status: translated from GH/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-327 <R0U>
A:Cross-references: EMBL:AC003974; NID:g2914688; PID:g2914702
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: E84736
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-327 <STO>
A:Cross-references: GB:AE002093; NID:g2914702; PID:AMC04492.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g32710; F24L7.15
A:Map position: 2
A:Introns: 193/2
C:Superfamily: Arabidopsis thaliana hypothetical protein F24L7.15

Query Match 28.7%, Score 60; DB 2; Length 327;
Best Local Similarity 33.3%, Pred. No. 12;
Matches 12; Conservative 5; Mismatches 19; Indels 0; Gaps 0;
QY 2 GISEADGKGNARGDKSIAGDIAQALGSGSTAI 37
DB 12 GAGAGAGGGGGGGGESSIALMDVVPSSSSSLGCV 47

Search completed: July 30, 2002, 15:35:50
Job time: 255 sec


```

Matches 10; Conservative 3; Mismatches 11; Indels 9; Gaps 2
OY      9 GKGGANRAGK---SIATGIDIAQ-----ALGSOSIATIGNDKT 42
        | ||| ||| ||||| : : ||| ||| : :
DB       94 GAGGLNRRANDPYISIAIGATAEAKPAVAVSGSIATGVNSV 136

RESULT      2
YADA_YEREN ID YADA_YEREN STANDARD: PRT; 455 AA.
P31489:
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Invasin precursor (Outer membrane adhesin).
GN YADA OR YOPA OR INVA OR YOPJ.
OS Yersinia enterocolitica.
OC Plasmid pYV.
CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
   Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6471/76 / SEROTYPE O:3;
RX MEDLINE=95020586; PubMed=7934875;
RA Tamm A., Jarkkanen A., Korhonen T.K., Kuusela P., Toivanen P.,
   Skurnik M.;
RA "Hydrophobic domains affect the collagen-binding specificity and
   surface polymerization as well as the virulence potential of the Yada
   protein of Yersinia enterocolitica.";
RL Mol. Microbiol. 10:995-1011(1993).
CC -!- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO
   PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE
   CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS
   PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL
   SURFACE.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on
   CC use by non-profit institutions as long as its content is in no way
   CC modified and this statement is not removed. Usage by and for commercial
   CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL, X13882; CAA32086.1; -.
KW PIR; S04912; S04912.
KW Plasmid; Virulence; Signal; Outer membrane.
FT SIGNAL 1 25
FT CHAIN 26 455 INVASIN
SQ SEQUENCE 455 AA; 47136 MW; AC12EF68C657DAC0 CRC64;

Query Match 30.4%; Score 63.5; DB 1; Length 455;
Best Local Similarity 44.2%; Pred.No.3.3;
Matches 19; Conservative 6; Mismatches 9; Indels 9; Gaps 2;

OY      9 GKGG--ANARDKSTAIKDIAQ-----ALGSOSIATIGNDKT 42
        ||| ||| ||| ||||| : : ||| ||| : :
DB       59 GAGGLNASAKGIHSIAIGATAEAAGAIVAAGAGSIATGVNSV 101

RESULT      3
MASZ_PSEFL ID MASZ_PSEFL STANDARD: PRT; 725 AA.
O05137:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Probable malate synthase G (EC 4.1.3.2).
GN GICH OR FC2 4.
OS Pseudomonas fluorescens.

```

```

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
RX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RA Strain-L6.5;
RA Artigueuve F.M., Delecu M., Villagines R., Danglot C.;
RT "A functional glyoxylate bypass is mandatory for utilization of
RT alkanes by Pseudomonas fluorescens."
RL Submitted (MAR-1997) to the EMBL/GenBank/DBS databases.
CC -| CATALYTIC ACTIVITY: L-malate + CoA -> acetyl-CoA + H(2)O +
CC glyoxylate.
CC -| PATHWAY: SECOND STEP IN GLYOXYLATE BYPASS, AN ALTERNATIVE TO THE
CC TRICARBOXYLIC ACID CYCLE (IN BACTERIA, FUNGI AND PLANTS).
CC -| SUBUNIT: MONOMER (BY SIMILARITY).
CC -| SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -| SIMILARITY: BELONGS TO THE MALATE SYNTHASE G FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use, by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL; Y11998; CAA72726.1; -
DR HSSP; P37330; 1D8C.
DR InterPro; IPR001465; Malate_synthase.
DR Pfam; PF01274; Malate_synthase; 1
KW Glyoxylate bypass; Tricarboxylic acid cycle; Lyase.
SQ SEQUENCE 725 AA; 78861 MW; 0C3326FE67C9B381 CRC64;

Query Match 29.9%; Score 62.5; DB 1; Length 725;
Best Local Similarity 40.7%; Pred. No. 6.8;
Matches 22; Conservative 2; Mismatches 17; Indels 13; Gaps 3;

OY 3 ISEADG--GKGGANRGDKSTAIG---DIAQALG-----SOSIAIGDNKIV 43
Db 148 ISEADGAERKGYNRGDKVIATFAFAELDENAPLSAGSHWSTGYKTIADGKLI 201

RESULT 4
PROA.CHLMU STANDARD; PRT; 377 AA.
ID PROA.CHLMU
AC Q9PUNA;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (Transcriptase
DE alpha chain) (RNA polymerase alpha subunit).
GN PF0A OR TC0794.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=833560;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-MOPN / NIGG;
RA MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,
RA White O., Hickey E.K., Peterson J., Overback T., Berry K., Bass S.,
RA Lithner K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwynn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MOpn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -| FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -| CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N).

```

CC -1- SUBUNIT: CONSISTS OF A SIGMA FACTOR AND THE RNAP CORE ENZYME WHICH
 CC IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, 1 BETA CHAIN AND 1
 CC OMEGA CHAIN (BY SIMILARITY).
 CC -1- DOMAIN: THE AMINO-TERMINAL PORTION IS INVOLVED IN THE ASSEMBLY OF
 CC CORE RNAP, WHEREAS THE C-TERMINAL IS INVOLVED IN INTERACTION WITH
 CC TRANSCRIPTIONAL REGULATORS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE ALPHA CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE002347; AAF39597.1; ALT_INIR.
 CC TIGR: TC0794;
 CC InterPro: IPR001700; RNA_POL_A_bac.
 CC Pfam: PF01000; RNA_POL_A_bac; 2.
 CC ProDom: PD001179; RNA_POL_A_bac; 1.
 CC Transferrase: Transcription; DNA-directed RNA polymerase;
 CC Complete proteome.
 CC SEQUENCE 377 AA: 4183 MW: 7DB20C7206BBD26 CRC64;
 SQ
 Query Match 29.2%; Score 61; DB 1; Length 377;
 Best Local Similarity 32.4%; Pred. No. 5.3;
 Matches 12; Conservative 8; Mismatches 17; Indels 0; Gaps 0;
 Oy 3 ISEADGKGKGANARCDKSAIGDIAAGSOSIAIGD 39
 Db 106 IODCEGGRASOKRATISIDASDLAAGGOKATLIGD 142
 ID 5
 TID7_DROME STANDARD; PRT; 278 AA;
 AC 027272; G9YX16;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription initiation factor TFIID 42 kDa subunit (TAII1-42)
 DE (TAII140) (p42) (Enhancer of yellow 1 protein).
 GN E(Y)1 OR TAF40 OR CG6474.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150630; PubMed=7545910;
 RA Kokubo T., Gong D.W., Wootton J.C., Horikoshi M., Roeder R.G.,
 RA Nakatani Y.,
 RT "Molecular cloning of Drosophila TFIID subunits."
 RL Nature 367:484-487(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=94037099; PubMed=8221891;
 RA Goodrich J.A., Hoey T., Thut C.J., Admon A., Tjian R.,
 RT "Drosophila TAF140 interacts with both a VP16 activation domain and
 RL the basal transcription factor TFIIB."
 RL Cell 75:519-530(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrita J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Bailey R.M., Basu A., Baxendale J., Bayraktiroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Padlos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doug L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mel M.-H., Ibegwam C.,
 RA Jastil M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Kethum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Markovlov G., Mishina N.V., Moberly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Splet E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J., Wu Q.A.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: TAFs ARE COMPONENTS OF THE TRANSCRIPTION FACTOR IID
 CC (TFIID) COMPLEX THAT ARE ESSENTIAL FOR MEDIATING REGULATION OF
 CC RNA POLYMERASE TRANSCRIPTION (BY SIMILARITY).
 CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -----
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -----
 CC -1- SIMILARITY: BELONGS TO THE TAF26 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U06458; AAC47347.1;
 CC EMBL: L29540; AAA28488.1;
 CC EMBL: AE003506; AAF48767.1;
 CC HSSP: P30129; ADPV.
 CC TRANSFAC: T02125;
 CC FlyBase: FBgn0000617; e(Y)1.
 CC InterPro: IPR000166; Histone.core.
 CC InterPro: IPR003162; TFIID-31.
 CC Pfam: PF02291; TFIID-31; 1.
 CC ProDom: PD011023; TFIID-31; 1.
 CC Transcription regulation; Nuclear protein.
 CC DOMAIN 190 197 POLY-THR.
 CC DOMAIN 202 209 POLY-GLY.
 CC DOMAIN 250 256 POLY-GLY.
 CC DOMAIN 267 274 POLY-GLU (ACIDIC).
 CC SEQUENCE 278 AA: 29314 MW: 0EA442C80467001F CRC64;
 SQ
 Query Match 28.7%; Score 60; DB 1; Length 278;
 Best Local Similarity 31.6%; Pred. No. 5.1;
 Matches 12; Conservative 8; Mismatches 18; Indels 0; Gaps 0;
 Oy 1 ISEADGKGKGANARCDKSAIGDIAAGSOSIAIG 38

DR PROSITE: PS00518: 2F_RING.1; 1.
 DR PROSITE: PS50089: 2F_RING.2; 1.
 KW Transcription regulation; Trans-acting factor; Activator; Zinc-finger;
 KW DNA-binding; Early protein; Repressor; Phosphorylation.
 FT ZN_FING 13 52 RING-TYPE.
 FT DOMAIN 284 331 ASP/GLU-RICH (ACIDIC).
 SO SEQUENCE 676 AA; 67701 MW; 99B0683C9BFC65D CRC64;

Query Match 27.5%; Score 57.5; DB 1; Length 676;
 Best Local Similarity 44.1%; Pred. No. 24;
 Matches 15; Conservative 4; Mismatches 14; Indels 1; Gaps 1;

OY 2 GISEADGKGGANRGD-KSIAIGDIAQALSSQS 34
 Db 120 GGSSEAGGAGGAEAGAGAGAGAGAGAGAGRA 153

RESULT 9
 YH77_ARCFU STANDARD; PRT; 245 AA.
 AC 028497;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein AF1777.
 GN AF1777.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 CC Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Peterson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kerevage A.R., Graham D.E., Kyrides N.C.,
 RA Fleischmann E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodok A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
 RA Sadov P.W., DAndrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:364-370(1997).
 CC -1- SIMILARITY: BELONGS TO THE UPF0135 (NIF) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AE000980; AAB89474.1; -;
 DR TIGR: AF1777; -;
 DR Interpro: IPR002678; DUF34.
 DR Pfam: PF01784; DUF34; 1.
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 245 AA; 26860 MW; AA516IDCF10B04D CRC64;

Db 125 EYGVKIGFSAKLEKATVGEIAEKIGPAMVLPGEERV 163

RESULT 10

ID RPOA_CHLTR STANDARD; PRT; 377 AA.

AC 046449; 084515;

DT 15-DEC-1998 (Rel. 37, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (Transcriptase

DE alpha-chain) (RNA polymerase alpha subunit).

GN RPOA OR CT507.

OS Chlamydia trachomatis.

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI_TaxID=813;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=L2/434/BU;

RX MEDLINE=95247702; PubMed=773029;

RA Gu L.J., Wenman W.M., Remacha M., Meuser R., Coffin J., Kaul R.;

RT "Chlamydia trachomatis RNA polymerase alpha subunit: sequence and

RT structural analysis.";

RL J. Bacteriol. 177:2594-2601(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=D/W-3/CX;

RX MEDLINE=99008093; PubMed=9784136;

RA Stephens R.S., Kalman S., Lammell C.J., Fan J., Marathe R., Aravind L.,

RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,

RA Davis R.W.;

RT "Genome sequence of an obligate intracellular pathogen of humans:

RT Chlamydia trachomatis.";

RL Science 282:754-759(1998).

CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION

CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS

CC SUBSTRATES (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +

CC (RNA)(N).

CC -1- SUBUNIT: CONSISTS OF A SIGMA FACTOR AND THE RNAP CORE ENZYME WHICH

CC IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, 1 BETA' CHAIN AND 1

CC OMEGA CHAIN (BY SIMILARITY).

CC -1- DOMAIN: THE AMINO-TERMINAL PORTION IS INVOLVED IN THE ASSEMBLY OF

CC CORE RNAP, WHEREAS THE C-TERMINAL IS INVOLVED IN INTERACTION WITH

CC TRANSCRIPTIONAL REGULATORS.

CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE ALPHA CHAIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

CC EMBL: L33834; AAA74989.1; -;

DR EMBL: AE001323; AAC68108.1; -;

DR HSSP: P00574; ICOO;

DR Sienna-2DPAGE: Q46449; -;

DR Interpro: IPR001700; RNA-pol_A_bac.

DR Pfam: PF01000; RNA-pol_A_bac; 1.

DR Pfam: PF03118; RNA-pol_A_CTD; 1.

DR Pfam: PF001179; RNA-pol_A_bac; 1.

KW Transferrase; Transcription; DNA-directed RNA polymerase;

KW Complete proteome.

FT VARIANT 30 30 G -> S (IN SEROVAR L2).

FT VARIANT 92 93 IV -> ML (IN SEROVAR L2).

FT VARIANT 124 124 I -> V (IN SEROVAR L2).

FT VARIANT 377 377 G -> A (IN SEROVAR L2).

FT VARIANT 295 295 R -> A (IN REF. 2).

FT CONFLICT 377 295 R -> A (IN REF. 2).

SO SEQUENCE 377 AA; 41824 MW; 09B55FBBC36804A CRC64;

OY 5 EADGKGGANRGDSIAIGDIAQALG-SQSIAIGDKKI 42

Query Match 27.0%; Score 56.5; DB 1; Length 245;
 Best Local Similarity 33.3%; Pred. No. 11;
 Matches 13; Conservative 12; Mismatches 13; Indels 1; Gaps 1;

Query Match 26.8%; Score 56; DB 1; Length 377;
 Best Local Similarity 29.7%; Pred. No. 20;
 Matches 11; Conservative 8; Mismatches 18; Indels 0; Gaps 0;

QY 3 ISEADGKGGKGANMGKSTAIQDAQLSGSQAICD 39
 DB 106 LQDEGGRCQKLTATISIDASDLAAGGKEVTLGD 142

RESULT 11
 NDDO_ALCXX STANDARD; PRT; 498 AA.

AC P94212;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE N-acyl-D-aspartate deacylase (EC 3.5.1.83) (N-acyl-D-aspartate
 deamidohydrolase).
 OS Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxydans).
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 OX Achromobacter.
 NCBI_TaxID=515;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A-6;
 RA Wakayama M., Watanabe E., Takenaka Y., Miyamoto Y., Tau Y., Sakai K.,
 Moriuchi M.;
 RT "Cloning, expression, and nucleotide sequence of the N-acyl-D-
 aspartate amidohydrolase gene from Alcaligenes xylosoxydans subsp.
 RT xylosoxydans A-6.";
 RL J. Ferment. Bioeng. 80:311-317(1995).
 RN [2]
 RP CHARACTERIZATION.
 RC STRAIN-A-6;
 RX MEDLINE=93372486; PubMed=7763985;
 RA Moriuchi M., Sakai K., Katsumo Y., Maki T., Wakayama M.;
 RT Purification and characterization of novel N-acyl-D-aspartate
 RT amidohydrolase from Alcaligenes xylosoxydans subsp. xylosoxydans
 RT A-6.";
 RL Biosci. Biotechnol. Biochem. 57:1145-1148(1993).
 CC -1- CATALYTIC ACTIVITY: N-acyl-D-aspartate + H(2)O = carboxylate + D-
 CC aspartate.
 CC -1- COFACTOR: ZINC.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE N-ACYL-D-AMINO-ACID DEACYLASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: D45919; BAA08350.1; -;
 KW Hydrolase; Zinc.
 SQ SEQUENCE 498 AA: 53584 MW; EAF8662891F9F1B CRC64;

Query Match 26.8%; Score 55.5; DB 1; Length 498;
 Best Local Similarity 40.0%; Pred. No. 29;
 Matches 14; Conservative 4; Mismatches 12; Indels 5; Gaps 1;

OY 2 GISEADGKGGKGAN-----ARGDKSTAIQDAQLG 31
 DB 20 GATLIDGGGGPARQGDLANRGGRIVYALGDFAHAPG 54

RESULT 12
 SK12_YEAST STANDARD; PRT; 1287 AA.
 AC P35207: 006047;
 DT 01-FEB-1994 (Rel. 28, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Antiviral protein SK12.
 GN SK12 OR YLR396C OR L8084.17.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93309467; PubMed=8321235;
 RA Widner W.R., Wickner R.B.;
 RT "Evidence that the SKI antiviral system of Saccharomyces cerevisiae
 RT acts by blocking expression of viral mRNA.";
 RL Mol. Cell. Biol. 13:4331-4341(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RX MEDLINE=97313267; PubMed=9169871;
 RA Johnston M., Hillier L., Riles L., Albertmann K., Andre B., Ansoorge W.,
 Benes V., Bruckner M., Delius H., Dubois E., Dusterhoft A.,
 Eutian K.D., Floeth M., Goffeau A., Hedling U., Heumann K.,
 Heuss-Weltzel D., Hilbert H., Hilger F., Kleine K., Kotter P.,
 Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Mostl D.,
 Muller-Auer S., Newtich U., Obermaier B., Piravandi E., Pohl T.M.,
 Portetelle D., Purnelle B., Rechemann S., Rieger M., Rinke M., Rose M.,
 Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
 Underwood A.P., Uristarazu L.A., Vandenbol M., Verhasselt P.,
 Vierendeels F., Voelt K., Volckaert G., Voss H., Wambutt R., Wedler E.,
 Wedler H., Zimmermann F.K., Zollner A., Hani J., Hohnselt J.D.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
 RL Nature 387:87-90(1997).
 RN [3]
 RP SEQUENCE OF 1-162 FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE=95116323; PubMed=7816623;
 RA Lygeron Z., Conesa C., Lesage P., Swanson R.N., Ruet A., Carlson M.,
 RA Sentenac A., Seraphin B.;
 RT "The yeast BDF1 gene encodes a transcription factor involved in the
 RT expression of a broad class of genes including snRNAs.";
 RL Nucleic Acids Res. 22:5332-5340(1994).
 CC -1- FUNCTION: REPRESENTS DSRNA VIRUS PROPAGATION BY SPECIFICALLY
 CC BLOCKING TRANSLATION OF VIRAL MRNAS. PERHAPS RECOGNIZING THE
 CC ABSENCE OF CAP OR POLY(A). ESSENTIAL FOR CELL GROWTH ONLY IN THE
 CC PRESENCE OF M1 REPLICON. SEEMS TO BE A HELICASE.
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
 CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. SK12 SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: L13469; AAA35049.1; -;
 DR EMBL: U19729; AAB82356.1; -;
 DR EMBL: Z18944; CAA79378.1; -;
 DR PIR: B48140; B48140.
 DR SGD: S0004390; SK12.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00270; DEAD_1.
 DR Pfam: PF00271; Helicase_C_1.
 DR SMART: SM00487; DEXDC_1.
 DR SMART: SM00490; HELICG_1.
 KW Antiviral; Translation regulation; RNA-binding; DNA-binding; Helicase;
 KW ATP-binding.
 FT NP_BIND 351 358 ATP (POTENTIAL).
 FT SITE 444 447 DEWV BOX.
 FT DOMAIN 556 577 RNA-BINDING RGG-BOX (BY SIMILARITY).
 FT DOMAIN 555 597 ARG/GLY-RICH.

FT CONFLICT 326 326 W -> C (IN REF. 1).
 FT CONFLICT 759 760 OM -> L (IN REF. 1).
 SQ SEQUENCE 1287 AA; 146058 MW; 7CDD36CFCD0F8C32 CRC64;

Query Match 25.8%; Score 54; DB 1; Length 1287;
 Best Local Similarity 40.9%; Pred. No. 1,1e+02;
 Matches 18; Conservative 5; Mismatches 17; Indels 4; Gaps 2;

OY 2 GISEADGKGAN--ARGDKSIAGDIAQALGSGST-ATGDNK 41
 DB 557 GSGTARGGSGNTRDGRGNGRSTRGANGSGSGAGCAIGSNK 600

RESULT 13
 MACS_BOVIN STANDARD: PRT; 331 AA.
 AC P12624;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Myristoylated alanine-rich C-kinase substrate (MARCKS) (ACAMP-81).
 GN MACS.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89282412; PubMed=2734111;
 RA Stumpo D.J., Graff J.M., Albert K.A., Greengard P., Blackshear P.J.;
 RT "Nucleotide sequence of a cDNA for the bovine myristoylated
 RT alanine-rich C kinase substrate (MARCKS).";
 RT Nucleic Acids Res. 17:3987-3988(1989).
 RL [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=89264553; PubMed=2726763;
 RA Stumpo D.J., Graff J.M., Albert K.A., Greengard P., Blackshear P.J.;
 RT "Molecular cloning, characterization, and expression of a cDNA
 RT encoding the '80- to 87-kDa' myristoylated alanine-rich C kinase
 RT substrate: a major cellular substrate for protein kinase C.";
 RT Proc. Natl. Acad. Sci. U.S.A. 86:4012-4016(1989).
 RL [3]
 RP PARTIAL SEQUENCE.
 RX MEDLINE=92171958; PubMed=1540183;
 RA Mizutani A., Tokumitsu H., Hidaka H.;
 RT "Acidic calmodulin binding protein, ACAMP-81, is MARCKS protein
 RT interacting with synapsin I.";
 RT Biochem. Biophys. Res. Commun. 182:1395-1401(1992).
 RL [4]
 RP PHOSPHORYLATION SITES.
 RX MEDLINE=89308594; PubMed=2473066;
 RA Graff J.M., Stumpo D.J., Blackshear P.J.;
 RT "Characterization of the phosphorylation sites in the chicken and
 RT bovine myristoylated alanine-rich C kinase substrate protein, a
 RT prominent cellular substrate for protein kinase C.";
 RT J. Biol. Chem. 264:11912-11919(1989).
 RL [5]
 RP PHOSPHORYLATION SITES, AND REVISIONS.
 RC TISSUE=Brain;
 RX MEDLINE=94308052; PubMed=8034575;
 RA Taniguchi H., Manenti S., Suzuki M., Titani K.;
 RT "Myristoylated alanine-rich C kinase substrate (MARCKS), a major
 RT proline-directed protein kinase(s), is an in vivo substrate of
 RT the post-translational modifications.";
 RT J. Biol. Chem. 269:18299-18302(1994).
 RL [6]
 RP REVERSIBLE ASSOCIATION WITH THE MEMBRANE.
 RX MEDLINE=9138951; PubMed=2034276;
 RA Thelen M., Rosen A., Nairn A.C., Aderem A.;
 RT "Regulation by phosphorylation of reversible association of a

RT myristoylated protein kinase C substrate with the plasma membrane.";
 RL Nature 351:320-322(1991).
 RN [7]
 RP ACTIN-FILAMENT CROSS-LINKING.
 RX MEDLINE=92220195; PubMed=1560845;
 RA Hartwig J.H., Thelen M., Rosen A., Janney P.A., Nairn A.C.,
 RA Aderem A.;
 RT "MARCKS is an actin filament crosslinking protein regulated by
 RT protein kinase C and calcium-calmodulin.";
 RL Nature 356:618-622(1992).
 CC -1- FUNCTION: MARCKS IS THE MOST PROMINENT CELLULAR SUBSTRATE FOR
 CC PROTEIN KINASE C. THIS PROTEIN BINDS CALMODULIN, ACTIN, AND
 CC SYNAPSIN. MARCKS IS A FILAMENTOUS (F) ACTIN CROSS-LINKING PROTEIN.
 CC -1- PTM: PHOSPHORYLATION BY PKC REPLACES MARCKS FROM THE MEMBRANE. IT
 CC ALSO INHIBITS THE F-ACTIN CROSS-LINKING ACTIVITY.
 CC -1- SIMILARITY: BELONGS TO THE MARCKS FAMILY.
 CC -1- CAUTION: REF.1 AND REF.2 SEQUENCES DIFFER FROM THAT SHOWN IN
 CC POSITIONS 140 TO 150 DUE TO A FRAMESHIFT.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL, M24638; AAA30635.1; ALT_FRAME.
 CC PIR, A32904; A32904.
 CC PIR, P50338; P50338.
 CC PIR, S08341; S08341.
 CC InterPro: IPR002101; MARCKS.
 CC Pfam: PF02063; MARCKS.1.
 CC PRINTS: PR00963; MARCKS.
 CC PROSITE: PS00826; MARCKS_1; 1.
 CC PROSITE: PS00827; MARCKS_2; 1.
 CC KW Phosphorylation; Myristate; Calmodulin-binding; Actin-binding;
 CC Membrane.
 CC KW INT. MET 0 0
 CC LIPID 1 1 MYRISTATE.
 CC DOMAIN 150 174 CALMODULIN-BINDING (PSD).
 CC FT MOD_RES 26 26 PHOSPHORYLATION.
 CC FT MOD_RES 45 45 PHOSPHORYLATION.
 CC FT MOD_RES 80 80 PHOSPHORYLATION.
 CC FT MOD_RES 99 99 PHOSPHORYLATION.
 CC FT MOD_RES 116 116 PHOSPHORYLATION.
 CC FT MOD_RES 133 133 PHOSPHORYLATION (BY PKC).
 CC FT MOD_RES 157 157 PHOSPHORYLATION (BY PKC).
 CC FT MOD_RES 161 161 PHOSPHORYLATION (BY PKC).
 CC FT MOD_RES 165 165 PHOSPHORYLATION (BY PKC).
 CC FT MOD_RES 168 168 PHOSPHORYLATION (BY PKC).
 CC SQ SEQUENCE 331 AA; 31450 MW; 0F77BF117EDB35CA CRC64;

Query Match 25.6%; Score 53.5; DB 1; Length 331;
 Best Local Similarity 42.9%; Pred. No. 33;
 Matches 15; Conservative 3; Mismatches 16; Indels 1; Gaps 1;

OY 5 EADGKGANANGDKSIAGDIAQALGSGSTAGD 39
 DB 181 EAEGA-AGASAGKDEAGSAGAAAGGAAPGE 214
 RESULT 14
 GLND_RHIME STANDARD: PRT; 442 AA.
 AC 087392;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Glutamate synthase large subunit-like protein.
 GN GLXD OR R00089 OR SMC02612.
 OS Rhizobium meliloti (Sinorhizobium meliloti).

OC Bacteria: Proteobacteria: alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae: Sinorhizobium.
 OX NCBI_TaxID=382;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Blaut J.,
 RA Bolstad P., Becker A., Boutry A., Gadieu E., Dreño S., Gloux S.,
 RA Godt T., Goiffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,
 RA Pohl T., Portetelle D., Puehler A., Furrelier B., Ramsperger U.,
 RA Renard C., Thebaud P., Vandendol M., Weidner S., Gallbert F.,
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 CC -I- SIMILARITY: TO GLUTAMATE SYNTHASES LARGE SUBUNITS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch)
 CC
 DR EMBL, AF055582; AAC62222.1; -
 DR EMBL, AL591782; CAC41476.1; -
 DR InterPro: IPR003009; FMN_enzyme.
 DR InterPro: IPR002932; glu_synthase.
 DR Pfam: PF01645; glu_synthase; 1.
 KW Oxidoreductase; Complete proteome.
 FT CONFLICT 322
 SO SEQUENCE 442 AA; 47376 MW; CCF0BB8EC2D614CE CRC64;

Query Match 25.6%; Score 53.5; DB 1; Length 442;
 Best Local Similarity 42.4%; Pred. No. 44;
 Matches 14; Conservative 8; Mismatches 8; Indels 3; Gaps 2;

OY 11 GGANRQD-KSIAGDTAALGSOS-TAIGDN 40
 Db 302 GGRSGADVAKALAGADAVAIGTAIVAIIGN 334

RESULT 15
 EEG_HAEIN
 ID ERG_HAEIN STANDARD; PRT; 699 AA.
 AC P43925;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Elongation factor G (EF-G).
 GN FUSA OR FUS OR HI0579.
 OS Haemophilus influenzae
 CC Bacteria: Proteobacteria; gamma subdivision; Pasteurellales;
 CC Haemophilus.
 OX NCBI_TaxID=727;
 RN
 RP SEQUENCE FROM N.A.
 RP STRAIN-RD / KM20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleisemann K.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavang A.R., Butt C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McCutney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Wetman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Unterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.T., Geoghagen N.S.M.,

| | |
|----------|--|
| RA | Gentile M.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., |
| RA | Venter J.C.; "Whole-genome random sequencing and assembly of Haemophilus |
| RT | influenzae Rd.," |
| RL | Science 269:446-512(1995). |
| CC | -1) FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLLOCATION |
| CC | OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE |
| CC | RIBOSOME. |
| CC | -1b SUBCELLULAR LOCATION: Cytoplasmic. |
| CC | -1c SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY. |
| CC | -1d EF-G/EF-2 SUBFAMILY. |
| CC | ----- |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - |
| CC | the European Bioinformatics Institute. There are no restrictions on its |
| CC | use by non-profit institutions as long as its content is in no way |
| CC | modified and this statement is not removed. Usage by and for commercial |
| CC | entities requires a license agreement (see http://www.isb.ch/announce/isb-sib.ch). |
| CC | or send an email to license@isb-sib.ch). |
| CC | ----- |
| DR | EMBL: U32739; AAC2237.1; . |
| DR | HSSP: P13551; IEFNM. |
| DR | TIGR: H10579; |
| DR | InterPro: IPK000640; EFG_C. |
| DR | InterPro: IPR000795; GTP_EFTU. |
| DR | InterPro: IPR004161; GTP_EFTU_D2. |
| DR | Pfam: PF006079; EFG_C; 1. |
| DR | Pfam: PF00009; GTP_EFTU; 1. |
| DR | Pfam: PF03144; GTP_EFTU_D2; 1. |
| DR | PRINTS: PR00315; ELONGATNCT. |
| DR | PROSITE: PS00301; EFACITOR_GTP; 1. |
| KW | Elongation factor; Protein biosynthesis; GTP-binding; |
| KW | Complete proteome. |
| FT | INIT_MET 0 |
| FT | NP_BIND 16 |
| FT | NP_BIND 87 |
| FT | NP_BIND 141 |
| FT | BY_SIMILARITY. |
| FT | GTP (BY SIMILARITY). |
| FT | GTP (BY SIMILARITY). |
| FT | GTP (BY SIMILARITY). |
| SEQUENCE | 699 AA; 77132 MW; FBBAD639C0F62801 CCK64; |

| | | | | |
|-----------------------|-------|-----------------------------|------|----------------------------------|
| Query Match | 25.6% | Score 53.5 | DB 1 | Length 659 |
| Best Local Similarity | 42.9% | Pred. No. 69 | | |
| Matches | 12 | Conservative | 5 | Mismatches 10; Indels 1; Gaps 1; |
| QY | 13 | ANARGD-KSTATGDAQALGSQSIAIGD | 39 | |
| | | | | |
| Db | 365 | ANKREIKKEVRADIAAIGLDVTTGD | 392 | |

Search completed: July, 30, 2002, 15:37:58
Job time: 312 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 30, 2002, 15:31:55 ; Search time 88.56 Seconds

(without alignments)
83.997 Million cell updates/sec

Title: US-09-813-214A-1

Perfect score: 209
Sequence: 1 IGISEADGKGGKGNARGDSK.....GDIQAQLSGSIAIGDNKIV 43

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.19:.*
1: sp.archaea:.*
2: sp.bacteria:.*
3: sp.fungi:.*
4: sp.human:.*
5: sp.invertebrate:.*
6: sp.mammal:.*
7: sp.mhc:.*
8: sp.organelle:.*
9: sp.phage:.*
10: sp.plant:.*
11: sp.rodent:.*
12: sp.virus:.*
13: sp.vertibrate:.*
14: sp.unclassified:.*
15: sp.virus:.*
16: sp.bacteriap:.*
17: sp.archaeap:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 77 | 36.8 | 1299 | 16 | Q9F3X6 |
| 2 | 72.5 | 34.7 | 1328 | 2 | Q9LAX0 |
| 3 | 70 | 33.5 | 1953 | 16 | Q98HJ2 |
| 4 | 68 | 32.5 | 2712 | 16 | Q9F3X5 |
| 5 | 67 | 32.1 | 1265 | 2 | Q9FPA0 |
| 6 | 67 | 32.1 | 1281 | 16 | Q9ZK07 |
| 7 | 66 | 31.6 | 1004 | 16 | Q9PD63 |
| 8 | 64.5 | 30.9 | 1190 | 16 | Q9PC04 |
| 9 | 64 | 30.6 | 1107 | 2 | Q9FZD8 |
| 10 | 63.5 | 30.4 | 454 | 2 | Q85267 |
| 11 | 63 | 30.1 | 261 | 11 | Q9EOT3 |
| 12 | 63 | 30.1 | 487 | 2 | Q9LA53 |
| 13 | 63 | 30.1 | 2059 | 16 | Q9PD50 |
| 14 | 62.5 | 29.9 | 436 | 2 | Q9S2C3 |
| 15 | 61 | 29.2 | 550 | 2 | Q9AVH8 |
| 16 | 61 | 29.2 | 567 | 2 | Q9AFP8 |

| | | | | | | |
|----|------|------|------|----|--------|----------------------|
| 17 | 61 | 29.2 | 615 | 16 | 005806 | 005806 mycobacteri |
| 18 | 61 | 29.2 | 694 | 16 | 053212 | 053212 mycobacteri |
| 19 | 60.5 | 28.9 | 608 | 5 | Q9VMM5 | Q9VMM5 drosophila |
| 20 | 60 | 28.7 | 253 | 5 | Q95R74 | Q95R74 drosophila |
| 21 | 60 | 28.7 | 286 | 10 | Q9ACM0 | Q9ACM0 arabidopsis |
| 22 | 60 | 28.7 | 289 | 10 | Q9FR83 | Q9FR83 arabidopsis |
| 23 | 60 | 28.7 | 327 | 10 | Q48846 | Q48846 arabidopsis |
| 24 | 60 | 28.7 | 449 | 5 | Q9V7C0 | Q9V7C0 glycoline max |
| 25 | 59.5 | 28.5 | 452 | 10 | Q9XE94 | Q9XE94 bacterioph |
| 26 | 59 | 28.2 | 511 | 9 | Q9MCT8 | Q9MCT8 bacterioph |
| 27 | 58.5 | 28.0 | 388 | 3 | Q9P557 | Q9P557 neurospora |
| 28 | 58.5 | 28.0 | 464 | 10 | Q9ZS34 | Q9ZS34 nicotiana t |
| 29 | 58 | 27.8 | 504 | 2 | Q9LAS6 | Q9LAS6 escherichia |
| 30 | 58 | 27.8 | 1176 | 12 | Q98587 | Q98587 paramicium |
| 31 | 57.5 | 27.5 | 403 | 4 | Q9UHR6 | Q9UHR6 homo sapien |
| 32 | 57.5 | 27.5 | 1660 | 16 | 053215 | 053215 mycobacteri |
| 33 | 57 | 27.3 | 422 | 2 | 056930 | 056930 yersinia en |
| 34 | 57 | 27.3 | 422 | 2 | Q93KR4 | Q93KR4 yersinia en |
| 35 | 57 | 27.3 | 737 | 2 | Q9AF04 | Q9AF04 frankia sp. |
| 36 | 57 | 27.3 | 1737 | 5 | Q9TW28 | Q9TW28 dictyostel |
| 37 | 57 | 27.3 | 2301 | 5 | Q95ZD0 | Q95ZD0 leishmania |
| 38 | 56.5 | 27.0 | 2342 | 5 | 001677 | 001677 bombyx mori |
| 39 | 56 | 26.8 | 369 | 2 | Q9AU00 | Q9AU00 bacillus sp |
| 40 | 56 | 26.8 | 552 | 5 | 021538 | 021538 caenorhabdi |
| 41 | 56 | 26.8 | 584 | 16 | 053575 | 053575 mycobacteri |
| 42 | 55.5 | 26.6 | 698 | 16 | Q9KU27 | Q9KU27 vibrio chol |
| 43 | 55 | 26.3 | 335 | 9 | Q9MBR0 | Q9MBR0 staphylococ |
| 44 | 55 | 26.3 | 335 | 5 | 061538 | 061538 drosophila |
| 45 | 55 | 26.3 | 399 | 16 | Q98WS4 | Q98WS4 rhizobium 1 |

ALIGNMENTS

RESULT 1
ID Q9F3X6 PRELIMINARY; PRT; 1299 AA.
AC Q9F3X6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MAPA PROTEIN (HSP).
GN MAPA OR HSF_2 OR PM1570.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins";
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RM [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145856; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AJ277635; CAC14202.1;
KW Complete proteome.
SQ SEQUENCE 1299 AA; 130963 MW; 8BCCE0EB66CDB428 CRC64;

Query Match 36.8%; Score 77; DB 16; Length 1299;
Best Local Similarity 45.7%; Pred. No. 1.3;
Matches 21; Conservative 7; Mismatches 12; Indels 6; Gaps 2;

QY 1 IGISEADGKGGK---GANKRDK--SIAIGDIAQLSGSIAIGDN 40
DB 105 IGFGATNDGTNVAIGAKSKSKASIASIAIGNAKLMDQALTAIGDN 154

RESULT 2
 09LAXO
 ID 09LAXO PRELIMINARY; PRT: 1328 AA.
 AC 09LAXO;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOHETICAL 127.4 KDA PROTEIN (FRAGMENT).
 OS Xanthomonas campestris pv. pelargonii.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xanthomonas.
 OX NCBI_TaxID=91612;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=XPEL-1;
 RA Kim J.F., Zumboff C.H., Beer S.V.;
 RT "An alanine-, glycine-, and serine-rich protein and a putative serine
 protease of Xanthomonas campestris pv. pelargonii."
 RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF083618; AAF63394.1; -
 KW Hypothetical protein.
 FT NON_TER
 SQ SEQUENCE 1328 AA; 127402 MW; ACE01A4C6548364C CRC64;

Query Match 34.7%; Score 72.5; DB 2; Length 1328;
 Best Local Similarity 43.5%; Pred. No. 4.7;
 Matches 20; Conservative 6; Mismatches 15; Indels 5; Gaps 2;
 QY 2 GISEADG--GKGC--ANARGDKSIAIGDIAQALGSGSIAIGDKNI 42
 Db 236 GLSTASGFLSSAGGYLSRASGASTAFGYRARESGSSIAVGFTAL 281

RESULT 3
 098HJ2 PRELIMINARY; PRT: 1953 AA.
 ID 098HJ2
 AC 098HJ2;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE MLL2848 PROTEIN.
 GN MLL2848.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpou S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti."
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP003000; BAB49874.1; -
 KW Complete proteome.
 SQ SEQUENCE 1953 AA; 184557 MW; 892BFA8B687B35E2 CRC64;

Query Match 33.5%; Score 70; DB 16; Length 1953;
 Best Local Similarity 35.4%; Pred. No. 14;
 Matches 17; Conservative 9; Mismatches 16; Indels 6; Gaps 1;
 QY 1 IGISEADGKGC-----ANARGDKSIAIGDIAQALGSGSIAIGDKNI 42
 Db 359 IGLGATAGCGVDTAIGKQANASQAQDAIAMGTSYAKASSQAQAIAGVNAV 406

RESULT 4
 09F3X5 PRELIMINARY; PRT: 2712 AA.
 ID 09F3X5
 AC 09F3X5;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE MAP PROTEIN (HSE).
 GN MAP OR HSE_1 OR PM0714.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
 RT "Evolutionary origins of the autotransporter proteins."
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida PM70."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 DR EMBL: AJ277636; CAC14203.1; -
 KW EMBL: AE006108; AAK02798.1; -
 SQ SEQUENCE 2712 AA; 276154 MW; 3F5579D6F32FFA3D CRC64;

Query Match 32.5%; Score 68; DB 16; Length 2712;
 Best Local Similarity 57.7%; Pred. No. 35;
 Matches 15; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 17 GIKSIAIGDIAQALGSGSIAIGDKNI 42
 Db 324 GESSVAIGDKAVSRGEASIAIGKNAI 349

RESULT 5
 09FDA0 PRELIMINARY; PRT: 1265 AA.
 ID 09FDA0
 AC 09FDA0;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PUTATIVE OUTER MEMBRANE PROTEIN XADA.
 GN XADA.
 OS Xanthomonas oryzae pv. oryzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xanthomonas.
 OX NCBI_TaxID=64187;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ray S.K., Rajeshwari R., Sonti R.V.;
 RT "A putative outer membrane protein from Xanthomonas oryzae pv. oryzae
 that is involved in virulence."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF288222; AAC01335.1; -
 SQ SEQUENCE 1265 AA; 119856 MW; 21762579B5EC70A2 CRC64;

Query Match 32.1%; Score 67; DB 2; Length 1265;
 Best Local Similarity 46.2%; Pred. No. 20;
 Matches 12; Conservative 8; Mismatches 6; Indels 0; Gaps 0;
 QY 13 ANARGDKSIAIGDIAQALGSGSIAIG 38
 Db 694 AATGVSAAVAIGFISKATGESVAVG 719

RESULT 6
 092K07 PRELIMINARY; PRT; 1291 AA.
 ID 092K07
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DE HYPOTHETICAL PROTEIN SMC01708.
 GN SMC01708.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21368234; PubMed=11474104;
 RA Galibert F., Fian T.M., Long S.R., Puehler A., Abola P., Ampe F.,
 Barloy-Hubler F., Barnett M.J., Becker A., Bolstad P., Bothe G.,
 Boutry M., Bowser L., Buhmester J., Cadieu E., Capela D., Chain P.,
 Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
 Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
 Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
 Kahn M.L., Kaiman S., Keating D.H., Kiss E., Komp C., Lelaire V.,
 Maury D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
 Rastberger U., Surzycki R., Thebault P., Vandenbol M.,
 Vorholter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
 RA "The composite genome of the legume symbiont Sinorhizobium meliloti.";
 RL Science 293:668-672(2001).
 DR EMBL; AL591783; CAC41895.1;
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 1291 AA; 127509 MW; 1B0F2A8CD1B13C CRC64;

Query Match 32.1%; Score 67; DB 16; Length 1291;
 Best Local Similarity 44.2%; Pred. No. 20;
 Matches 19; Conservative 3; Mismatches 13; Indels 8; Gaps 1;
 Oy 4 SEADGKGG-----ANAGDKSIAIGDIAQALGSOSIAIG 38
 Db 88 NEADPGASDAIAIGTDIAQANDRSIAIGRONAGNEGSIGIG 130
 RESULT 7
 09PD63 PRELIMINARY; PRT; 1004 AA.
 ID 09PD63
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DE SURFACE-EXPOSED OUTER MEMBRANE PROTEIN.
 GN XP1516.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 CC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
 Coutinho N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 Garlier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 Ho P.L., Hohnselt J.D., Junqueira M.L., Kempner E.L., Kitajima J.P.,
 Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Matsukuma A.Y.,
 Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 Peixoto B.R., Pereira G.A.G., Rodrigues V., de Rosa A.J.M.,
 de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 da Silva J.F., da Silveira M.L.Z., Siqueira W.J., de Souza A.A.,
 de Souza A.P., Terezi M.F., Truffi D., Tsai S.M., Tsuchiko M.H.,
 Vaitava H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,

RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Matsukuma A.Y.,
 Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 Queiroz R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 da Silva J.F., da Silveira M.L.Z., Siqueira W.J., de Souza A.A.,
 de Souza A.P., Terezi M.F., Truffi D., Tsai S.M., Tsuchiko M.H.,
 Vaitava H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 Zago M.A., Zatz M., Zeldin J., Zetser J.C.;
 RA "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RT Nature 406:151-159(2000).
 RL EMBL; AE003981; AAF84325.1;
 DR EMBL; AE003981; AAF84325.1;
 KW Complete proteome.
 SO SEQUENCE 1004 AA; 98283 MW; 0BFF04D0F048B11 CRC64;

Query Match 31.6%; Score 66; DB 16; Length 1004;
 Best Local Similarity 50.0%; Pred. No. 20;
 Matches 15; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
 Oy 9 GKGANAKGKSIAGDIAQALGSOSIAIG 38
 Db 576 GSGVASAIGKDSIATGASQAQVDSVALG 605
 RESULT 8
 09PC04 PRELIMINARY; PRT; 1190 AA.
 ID 09PC04
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DE SURFACE PROTEIN.
 GN XP1981.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 CC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
 Coutinho N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 Garlier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 Ho P.L., Hohnselt J.D., Junqueira M.L., Kempner E.L., Kitajima J.P.,
 Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Matsukuma A.Y.,
 Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 Peixoto B.R., Pereira G.A.G., Rodrigues V., de Rosa A.J.M.,
 de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 da Silva J.F., da Silveira M.L.Z., Siqueira W.J., de Souza A.A.,
 de Souza A.P., Terezi M.F., Truffi D., Tsai S.M., Tsuchiko M.H.,
 Vaitava H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,

RA Zago M.A., Zatz M., Meldanis J., Setunai J.C.;
RT "The genome sequence of the plant pathogen *Xylella fastidiosa*."
RL Nature 406:151-159(2000).
DR EMBL; AE004017; AAF84783.1; -.
KW Complete proteome.
SQ SEQUENCE 1190 AA; 118446 MW; 756741B0C8D787CC CRC64;

| | | | | |
|-----------------------|-----------------|----------------|-----------|--------------|
| Query Match | 30.9%; | Score 64.5; | DB 16; | Length 1190; |
| Best Local Similarity | 44.4%; | Pred. No. 36; | | |
| Matches 16; | Conservative 7; | Mismatches 12; | Indels 1; | Gaps 1 |

QY 4 SEADGCKGGAN-AREDKSIAIGDIAQALGSGSIAIG 38
1: 11 11 1: 1111 1: 1: 11
Db 191 SKTDGNTSGATVAQGLRSIAIGTTARSQSQDAISIG 226

| | | |
|--------|--------|--------------|
| RESULT | 9 | |
| Q9F2D8 | | |
| ID | Q9F2D8 | PRELIMINARY; |
| AC | Q9F2D8 | PRT; 1107 AA |

DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update),
DE SAPR PROTEIN

GN
SAPB.
OS
Salmonella typhi.
OC
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC
Salmonella.

| | |
|----|--------------------|
| 0A | NCBI_taxid=801; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RC | STRAIN=CT18; |

DR EMBL; AJ277623; CAC14217.1; -.

Query Match 30.68; Score 64; DB 2; Length 1107;

| | | | | | | | | | |
|---------|-----|---|----|------------|-----|--------|----|------|----|
| Matches | 15; | Conservative | 7; | Mismatches | 15; | Indels | 2; | Gaps | 1; |
| OY | 4 | SEADGCKGAMNRGDKSIAIGDIAQALGSSQSTAIGDKTI | 42 | | | | | | |

Db 879 TINTD--ADANAGADSVATGSGSIAAENSVALGTNSV 915

| | | | | | |
|----|-------------|--------------|------|----------|----|
| ID | 085267 | PRELIMINARY; | PRT; | 454 | AA |
| AC | 085267; | | | | |
| DT | 01-NOV-1998 | (TREMblrel) | 08 | Created) | |

DE 01-MAY-1999 (TREMBLERel. 10, last annotation update),
ADHESIN.
GN YADA.

03 *Legionella pneumophila* serogroup 1.
06 plasmid pVY, and plasmid pVY227.
0C Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae.
0C Yersinia.

CA. MOBL-100-030;
RN
RP SEQUENCE FROM N.A.
RA Iriarte M., Kerbouch C., Lambermont I., Cornelis G.R.;
[1]

Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
[2]

RA Iriarte M., Kebbourch C., Cornelis G.R.;
RT "Yada and ORF291 of Yersinia enterocolitica O:3.",

Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases

| RP | SEQUENCE FROM N.A. |
|----|--------------------|
| RC | STRAIN=W22703; |

RT "Detailed genetic map of the pV227 plasmid of *Yersinia enterocolitica* serotype O:3,";

| | |
|----|-------------------------------|
| DR | EMBL; AF056092; AAC33679.1; - |
| DR | EMBL; AF102990; AAD16868.1; - |
| KM | Plasmid. |

| Query Match | Score | DB 2 | Length |
|------------------------|-------|------|--------|
| Post local cultivation | 30.4% | 63.5 | 44 |
| Post local cultivation | 44.3% | 63.5 | 44 |
| Post local cultivation | 44.3% | 63.5 | 44 |

| Matches | Conservative | Mismatches | Indels | Gaps |
|---------|--------------|------------|--------|------|
| 19; | 5; | 10; | 9; | 2 |

QY 9 GKGAN--ARGDKSIATGDIQA-----ALGSOSIATGDNKI 42

```

D6      59  GAGCINCBAKGIHSIATGATMEAAAGAAVAAGAGSIAGVNSV 101
RESULT  11

```

| | | |
|---------|---------------------------------------|-------------|
| Q9EQT3 | | |
| ID | PRELIMINARY; | PRT; 261 AA |
| Q9EQT3; | | |
| DT | 01-MAR-2001 (TREMBlere), 16. Created) | |

01-DEC-2001 (TREMBLE). 19, last annotation update)
GTP-BINDING PROTEIN LINE 1 (WRCH-1).
WRCH1 OR MG28K.

mus musculus (mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
 NCBI_TaxID=10090;

SEQUENCE FROM N.A.
Daigo Y., Takayama
"Isolation, mapping

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

| | |
|------------|---|
| RA | Tao W., Pennica D., Xu L., Kalejta R.F., Levine A.J.; |
| RT | "Wrch-1, a novel member of the Rho gene family that is regulated by |
| RT | Wnt-1. " |
| Genes Dev. | 15:1706-1007(2001). |

CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY
EMBL; AB051827; BAB18639.1; -.
DR EMBL; AF378088; AAK83341.1; -.
R HSSD. P21181. 1A.M4

DR MGD; MG1:1916831; wrchl.
DR InterPro; IPR001806; Ras_trnsfrmng.
DR InterPro; IPR003578; Rho.
DR PRINTS: PR00449; RASTRNSFRMG

DR SMART; SMOU1/4; RHU; 1.
KW GTP-binding; Lipoprotein; Prenylation.
SQ SEQUENCE 261 AA; 28353 MW; 8EC4250071E75B14 CRC64

| | | | | |
|--------------------------|--------|----------------|-----------|-------------|
| Query Match | 30.1%; | Score 63; | DB 11; | Length 261; |
| Best Local Similarity | 38.9%; | Pred. No. 9.9; | | |
| Matches 14; Conservative | 6; | Mismatches 14; | Indels 2; | Gaps 1; |

QY 2 GISEADGGKGANARGDKSLAIGDIAQLGSSQSLAI 37
1:1 1 111 11: :11 1:1 1: :
Db 37 GYSGGRGRAGGAEGRGYKCYLVGD--GAVGKTSLV 70

| | |
|--------|----|
| RESULT | 12 |
| Q9LA53 | |

ID Q9LA53 PRELIMINARY; PRT; 487 AA.
 AC Q9LA53;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE IMMUNOGLOBULIN-BINDING PROTEIN E1BE.
 GN E1BE.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ECOR-9;
 RX MEDLINE=20187497; PubMed=10722621;
 RA Sandt C.H., Hill C.W.;
 RT "Four different genes responsible for nonimmune immunoglobulin-binding activities within a single strain of Escherichia coli."
 RL Infect. Immun. 68:2205-2214(2000).
 DR EMBL: AF151676; AAF63045.1; -
 SQ SEQUENCE 487 AA; 51605 MW; D959C67EF7C21CF3 CRC64;
 Query Match 30.1%; Score 63; DB 2; Length 487;
 Best Local Similarity 33.3%; Pred. No. 20;
 Matches 19; Conservative 7; Mismatches 11; Indels 20; Gaps 2;
 OY 2 GISEADGKG-----GANARDKSIAIGDIAQ-----ALGSGSIAIG 38
 DB 138 GSVSDGKGVSLGSDAKSGEFSVAVNGAKATEKASTVSGWSAALGROSVALG 194
 RESULT 13
 Q9PD50 PRELIMINARY; PRT; 2059 AA.
 AC Q9PD50;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE SURFACE PROTEIN.
 GN Xf1329.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 CC Xylella.
 OX NCBI_TaxID=23711;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M., Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S., Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S., Bueno M.R.P., Camargo A., Camargo L.E.A., Carraro D.M., Carter H., Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M., Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H., Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A., Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R., Gariet M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A., Ho P.L., Hohnselt J.D., Junqueira M.U., Kemper E.L., Kitajima J.P., Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C., Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A., Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L., Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y., Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B., Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S., Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A., de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Pesquero J.B., de Paula B.R., Pereira G.A.G., Pereira H.A.Jr., Pasquero J.B., Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.U.M., de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.B., da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A.Jr., da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A., de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,

RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L., Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa."
 RL Nature 406:151-159(2000).
 DR EMBL: AE003982; AAF8438.1; -
 KW Complete proteome.
 SQ SEQUENCE 2059 AA; 204035 MW; EBA6A71B5D24E10 CRC64;
 Query Match 30.1%; Score 63; DB 16; Length 2059;
 Best Local Similarity 36.7%; Pred. No. 1e+02;
 Matches 18; Conservative 4; Mismatches 19; Indels 8; Gaps 1;
 OY 2 GISEADGKG-----GANARDKSIAIGDIAQALGSGSIAIGDNKI 42
 DB 1731 GNYDGDATGSKAIAAGVGTQASGEAAVSGGAAGKSGSTAIARRAI 1779
 RESULT 14
 Q9S2C3 PRELIMINARY; PRT; 436 AA.
 AC Q9S2C3;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PROBABLE SOLUTE-BINDING PROTEIN.
 GN SGP85.18.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Collum J., Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL110470; CAB54175.1; -
 DR InterPro: IPR000567; SBP_Dbc_1.
 DR Pfam: PF01547; SBP_bacterial_1;
 SQ SEQUENCE 436 AA; 46593 MW; 55815BEDDA6C4C04 CRC64;
 Query Match 29.9%; Score 62.5; DB 2; Length 436;
 Best Local Similarity 35.8%; Pred. No. 20;
 Matches 19; Conservative 5; Mismatches 16; Indels 13; Gaps 2;
 OY 2 GISEADGKG-----KCGA-----NARGDKSIAIGDIAQALGSGSIAIGDNK 41
 DB 283 GIGADGDGDFVSSGLGTMNPPYDGKGDPSNVAVNPAYLISSTKATBEK 335
 RESULT 15
 Q9AJV8 PRELIMINARY; PRT; 550 AA.
 AC Q9AJV8;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE U... PROBABLE PERIPLASMIC UDP-SUGAR HYDROLASE.

SN [USHA](#).
OS *Shigella flexneri*.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC *Shigella*.
OX NCBI_TaxID=623;
RN [1]
RP
RC STRAIN=M90T;
RX MEDLINE=20566792; Pubmed=1111511;
RA Buchrieser C., Glaeser P., Rusniok C., Nedjari H., d'Hauterville H.,
RA Kunst F., Sansonetti P., Paros C.;
RT "The virulence plasmid pWR100 and the repertoire of proteins secreted
RL by the type III secretion apparatus of *Shigella flexneri*."
RL Mol. Microbiol. 38:760-771(2000).
DR EMBL; AL391753; CAC05838.1; -.
DR HSSD; P07024; 2USH.
DR InterPro: IPR002224; 5_nucleotidase.
DR InterPro: IPR000934; Ser_thr_phosphase.
DR Pfam; PF01009; 5_nucleotidase; 1.
DR PROSITE; PS00785; 5_NUCLEOTIDASE; 1.
DR PROSITE; PS00786; 5_NUCLEOTIDASE_2; 1.
KW Hydrolase.
SQ SEQUENCE 550 AA; 61199 MW; ABAFE6B2165DD3A83 CRC64;

| | | | | | |
|----|---|--------|--------------|---------|----------------------------------|
| | Query Match | 29.2% | Score 61: | DB 2; | Length 550; |
| | Best Local Similarity | 37.2%; | Pred. | No. 39; | |
| | Matches | 16; | Conservative | 6; | Mismatches 13; Indels 8; Gaps 2. |
| Qy | 7 DGGKGNARCD-----KSTATIDTDAALGSSGSIA-----IGDNKK | 41 | | | |
| Dd | 222 DNKGSGSNAPFGDEMAKSLPTOSLMTLVGGHSHQAFLCYCAASDNNK | 264 | | | |

Search completed: July 30, 2002, 15:37:26
Job time: 331 sec

[illegible]

CC This is a peptide fragment of a novel outer membrane protein-106 (OMP106)
 CC polypeptide. The OMP106 is an outer membrane polypeptide of Moraxella
 CC catarhalis, an haemagglutinating cultivar. The peptide fragment can
 CC specifically bind to an antibody that binds the OMP106 polypeptide. The
 CC antibody is a cytotoxic antibody which mediates complement killing of
 CC M. catarhalis. The OMP106 polypeptide, and its peptide fragments can be
 CC used in vaccines and antigenic compositions. They can also be used for
 CC producing an immune response in an animal against M. catarhalis.
 CC
 SQ Sequence 8 AA;

Query Match 100.0%; Score 41; DB 18; Length 8;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVLGKK 8
 DB 1 gtlvlgkk 8

RESULT 2

AAE00704
 ID AAE00704 standard; peptide: 8 AA.

AC AAE00704;

DT 02-JUL-2001 (first entry)

DE Internal peptide of M. catarhalis outer membrane protein-106 (OMP106).

KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 bacterial infection; immunogen; cytotoxic; antibiotic;

KW passive immunisation.

OS Moraxella catarhalis.

PN US6214981-B1.

PD 10-APR-2001.

PE 12-NOV-1997; 97US-0968685.

PR 03-MAY-1996; 96US-0642712.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Tucker K, Plosila L, Tillman UF;

DR WPI: 2001-281002/29.

XX Novel nucleotide sequences encoding Moraxella catarhalis outer
 PT membrane protein-106 polypeptide, useful for diagnosis of bacterial
 PT infections and as vaccine against Moraxella catarhalis infection of
 PT mammals -

PS Example: Column 31; 49pp; English.

CC The present sequence is internal peptide of haemagglutinating
 CC Moraxella catarhalis outer membrane protein-106 (OMP106). The OMP106 is
 CC used as a therapeutic and prophylactic vaccine against M. catarhalis
 CC infections of mammals. It is used for diagnosis of bacterial infections
 CC and as reagents for clinical or medical diagnosis of M. catarhalis
 CC infections and for scientific research on the properties of
 CC pathogenicity, virulence and infectivity of M. catarhalis. It is also
 CC used as a probe to identify the presence of M. catarhalis in biological
 CC specimens and to identify other bacteria that encode a polypeptide
 CC related to M. catarhalis OMP106. OMP106-derived polypeptides are used
 CC as ligands to detect antibodies elicited in response to M. catarhalis
 CC infections and also as immunogens for inducing M. catarhalis-specific
 CC antibodies which are useful in immunoassays to detect M. catarhalis in
 CC biological specimens. Cytotoxic antibodies are useful in passive
 CC immunisations against M. catarhalis.

XX Sequence 8 AA;

SQ Query Match 100.0%; Score 41; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVLGKK 8
 DB 1 gtlvlgkk 8

RESULT 3

AAE00817
 ID AAE00817 standard; peptide: 7 AA.

AC AAE00817;

DT 02-JUL-2001 (first entry)

DE Peptide related to the invention.

KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 bacterial infection; immunogen; cytotoxic; antibiotic;

KW passive immunisation.

OS Unidentified.

PN US6214981-B1.

PD 10-APR-2001.

PE 12-NOV-1997; 97US-0968685.

PR 03-MAY-1996; 96US-0642712.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Tucker K, Plosila L, Tillman UF;

DR WPI: 2001-281002/29.

XX Novel nucleotide sequences encoding Moraxella catarhalis outer
 PT membrane protein-106 polypeptide, useful for diagnosis of bacterial
 PT infections and as vaccine against Moraxella catarhalis infection of
 PT mammals -

PS Disclosure: Column 63; 49pp; English.

CC The present sequence is a peptide which is related to the invention.
 CC The present invention relates to haemagglutinating
 CC Moraxella catarhalis outer membrane protein-106 (OMP106). The OMP106 is
 CC used as a therapeutic and prophylactic vaccine against M. catarhalis
 CC infections of mammals. It is used for diagnosis of bacterial infections
 CC and as reagents for clinical or medical diagnosis of M. catarhalis
 CC infections and for scientific research on the properties of
 CC pathogenicity, virulence and infectivity of M. catarhalis. It is also
 CC used as a probe to identify the presence of M. catarhalis in biological
 CC specimens and to identify other bacteria that encode a polypeptide
 CC related to M. catarhalis OMP106. OMP106-derived polypeptides are used
 CC as ligands to detect antibodies elicited in response to M. catarhalis
 CC infections and also as immunogens for inducing M. catarhalis-specific
 CC antibodies which are useful in immunoassays to detect M. catarhalis in
 CC biological specimens. Cytotoxic antibodies are useful in passive
 CC immunisations against M. catarhalis.
 CC Note: No other information is provided for this sequence in the
 CC specification.

SX Sequence 7 AA;

Query Match 87.8%; Score 36; DB 22; Length 7;

Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVLGK 7
111111
Db 1 gtlvgk 7

RESULT 4

AAW04505 standard; Protein; 1992 AA.

AAW04505;

25-JAN-1997 (first entry)

Moraxella 200 kDa outer membrane protein.

Outer membrane protein; OMP; immunogen; vaccine; otitis media; diagnosis.

Mycobacterium catarrhalis strain 4223.

MO9634960-A1

07-NOV-1996.

29-APR-1996; 96WO-CA00264.

26-MAR-1996; 96US-0621944.

01-MAY-1995; 95US-0431718.

07-JUN-1995; 95US-0478370.

(CONN-) CONNAUGHT LAB LTD.

Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;

WPI; 1996-506162/50.

N-PSDB; AAT38740.

Claim 14; Fig 6; 109pp; English.

An approx. 200 kDa outer membrane protein (AAW04505) can be

isolated from Moraxella catarrhalis otitis media strain 4223

by electroporation, or expressed from a gene (see also AAT38740)

obtd. from a strain 4223 genomic library. Natural or recombinant

outer membrane protein is useful as an immunogen to protect

against infection by Moraxella, esp. M. catarrhalis. It can

also be used to detect antibodies, esp. for differential diagnosis

between bacteria that cause similar symptoms; and also useful as

a carrier for other antigens and used to raise antitumour

antibodies for conjugation to therapeutic agents.

Sequence 1992 AA;

OY 1 GTVLGK 7

111111

Db 1604 gtlvgk 1610

RESULT 5

AAW04505 standard; Protein; 1992 AA.

XX 24-APR-2001 (first entry)

DE M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein SEQ ID NO:3.

XX Moraxella catarrhalis strain 4223; major outer membrane protein;

KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

OTITIS media; detection.

OS Moraxella catarrhalis.

PN WO200107619-A1.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-CA00870.

PR 27-JUL-1999; 99US-0361619.

PA (CONN-) CONNAUGHT LAB LTD.

PI Loosmore SM, Sasaki K, Yang Y, Klein MH;

DR WPI; 2001-159722/16.

DR N-PSDB; AAF59100, AAF59101.

PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

PT useful in protective vaccines and for diagnosis

PS Example 3; Fig 2A-W; 247pp; English.

XX The present invention describes an isolated and purified nucleic acid (I)

CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.

CC The 200 kDa outer membrane protein (II) has antibacterial activity and

CC can be used in vaccines. (II), and its truncated versions, are used as

CC immunogenic compositions and vaccines to protect against M. catarrhalis

CC infections, particularly otitis media in humans. (II) is also used as

CC antigen in immunoassays for detecting specific antibodies (Ab), and to

CC generate Ab. (I) are used for recombinant production of (II) and its

CC fragments are used as probes for identifying/cloning 200 kDa protein

CC genes from other strains, and for diagnostic detection of M. catarrhalis.

CC (I) makes possible production of large amount of recombinant immunogens.

CC Expression of truncated versions of (II) reduces toxicity of the protein

CC towards the Escherichia coli host. The present sequence represents the

CC M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein, which is

CC used in the exemplification of the present invention.

XX Sequence 1992 AA;

OY 1 GTVLGK 7

111111

Db 1604 gtlvgk 1610

RESULT 6

AAW04505 standard; Protein; 1992 AA.

XX 24-APR-2001 (first entry)

DE M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein SEQ ID NO:3.

XX Moraxella catarrhalis strain 4223; major outer membrane protein;

KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

OTITIS media; detection.

OS Moraxella catarrhalis.

PN WO200107619-A1.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-CA00870.

PR 27-JUL-1999; 99US-0361619.

PA (CONN-) CONNAUGHT LAB LTD.

PI Loosmore SM, Sasaki K, Yang Y, Klein MH;

DR WPI; 2001-159722/16.

DR N-PSDB; AAF59100, AAF59101.

PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

PT useful in protective vaccines and for diagnosis

PS Example 3; Fig 2A-W; 247pp; English.

XX The present invention describes an isolated and purified nucleic acid (I)

CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.

CC The 200 kDa outer membrane protein (II) has antibacterial activity and

CC can be used in vaccines. (II), and its truncated versions, are used as

CC immunogenic compositions and vaccines to protect against M. catarrhalis

CC infections, particularly otitis media in humans. (II) is also used as

CC antigen in immunoassays for detecting specific antibodies (Ab), and to

CC generate Ab. (I) are used for recombinant production of (II) and its

CC fragments are used as probes for identifying/cloning 200 kDa protein

CC genes from other strains, and for diagnostic detection of M. catarrhalis.

CC (I) makes possible production of large amount of recombinant immunogens.

CC Expression of truncated versions of (II) reduces toxicity of the protein

CC towards the Escherichia coli host. The present sequence represents the

CC M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein, which is

CC used in the exemplification of the present invention.

XX Sequence 1992 AA;

OY 1 GTVLGK 7

111111

Db 1604 gtlvgk 1610

RESULT 6

AAW04505 standard; Protein; 1992 AA.

OS Moraxella catarrhalis.
 XX WO200107619-A1.
 XX 01-FEB-2001.
 XX 26-JUL-2000; 2000WO-CA00870.
 XX 27-JUL-1999; 99US-0361619.
 PR (CONN-) CONNNAUGHT LAB LTD.
 PA Loosmore SM, Sasaki K, Yang Y, Klein MH;
 PI WPI: 2001-159722/16.
 DR N-PSDB; AAF59106.
 XX
 PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
 PT useful in protective vaccines and for diagnosis
 PS Claim 1: Fig 8A-V; 247pp; English.
 CC The present invention describes an isolated and purified nucleic acid (I)
 CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
 CC The 200 kDa outer membrane protein (II) has antibacterial activity and
 CC can be used in vaccines. (II), and its truncated versions, are used as
 CC immunogenic compositions and vaccines to protect against M. catarrhalis
 CC infections, particularly otitis media in humans. (II) is also used as
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to
 CC generate Ab. (I) are used for recombinant production of (II) and its
 CC fragments are used as probes for identifying/cloning 200 kDa protein
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.
 CC (I) makes possible production of large amount of recombinant immunogens.
 CC Expression of truncated versions of (II) reduces toxicity of the protein
 CC towards the Escherichia coli host. The present sequence represents the
 CC M. catarrhalis M56 200kDa protein in pKS348, which is given in the
 CC exemplification of the present invention.
 CC
 SQ Sequence 1992 AA;

Query Match 87.8%; Score 36; DB 22; Length 1992;
 Best Local Similarity 100.0%; Pred. No. 4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVLGK 7
 |||||
 DB 1604 gtlvgk 1610

RESULT 7
 AAB69134
 ID AAB69134 standard; Protein; 2047 AA.
 XX
 AC AAB69134;

DT 24-APR-2001 (first entry)

DE M. catarrhalis strain 4223 genomic 200kDa protein SEQ ID NO:7.

KW Moraxella catarrhalis strain 4223; major outer membrane protein;
 KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
 KM Otitis media; detection.

OS Moraxella catarrhalis.

XX WO200107619-A1.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX
 PA (CONN-) CONNNAUGHT LAB LTD.
 PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
 DR WPI: 2001-159722/16.
 DR N-PSDB; AAF59102, AAF59103.
 XX
 PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
 PT useful in protective vaccines and for diagnosis
 PS Claim 1: Fig 3A-W; 247pp; English.
 CC The present invention describes an isolated and purified nucleic acid (I)
 CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
 CC The 200 kDa outer membrane protein (II) has antibacterial activity and
 CC can be used in vaccines. (II), and its truncated versions, are used as
 CC immunogenic compositions and vaccines to protect against M. catarrhalis
 CC infections, particularly otitis media in humans. (II) is also used as
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to
 CC generate Ab. (I) are used for recombinant production of (II) and its
 CC fragments are used as probes for identifying/cloning 200 kDa protein
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.
 CC (I) makes possible production of large amount of recombinant immunogens.
 CC Expression of truncated versions of (II) reduces toxicity of the protein
 CC towards the Escherichia coli host. The present sequence represents the
 CC M. catarrhalis strain 4223 genomic 200kDa protein, which is given in the
 CC exemplification of the present invention.
 CC
 SQ Sequence 2047 AA;

Query Match 87.8%; Score 36; DB 22; Length 2047;
 Best Local Similarity 100.0%; Pred. No. 4.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVLGK 7
 |||||
 DB 1659 gtlvgk 1665

RESULT 8
 ID AAB69135
 ID AAB69135 standard; Protein; 2053 AA.
 XX
 AC AAB69135;

DT 24-APR-2001 (first entry)

DE M. catarrhalis strain Q8 200kDa protein SEQ ID NO:9.

KW Moraxella catarrhalis strain Q8; major outer membrane protein;
 KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
 KM Otitis media; detection.

OS Moraxella catarrhalis.

XX WO200107619-A1.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 99US-0361619.

XX (CONN-) CONNNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI: 2001-159722/16.

XX N-PSDB; AAF59104.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
 PT

PT useful in protective vaccines and for diagnosis

XX Claim 1; Fig 4A-V; 247bp; English.

XX The present invention describes an isolated and purified nucleic acid (I)

XX that encodes a 200 kDa outer membrane protein of *Moraxella catarrhalis*.

XX The 200 kDa outer membrane protein (II) has antibacterial activity and

XX can be used in vaccines. (II), and its truncated versions, are used as

XX immunogenic compositions and vaccines to protect against *M. catarrhalis*

XX infections, particularly otitis media in humans. (II) is also used as

XX antigen in immunoassays for detecting specific antibodies (Ab), and to

XX generate Ab. (I) are used for recombinant production of (II) and its

XX fragments are used as probes for identifying/cloning 200 kDa protein

XX genes from other strains, and for diagnostic detection of *M. catarrhalis*.

XX (I) makes possible production of large amount of recombinant immunogens.

XX Expression of truncated versions of (II) reduces toxicity of the protein

XX towards the *Escherichia coli* host. The present sequence represents the

XX exemplification of the present invention.

SO Sequence 2053 AA:

Query Match 87.8%; Score 36; DB 22; Length 2053;

Best Local Similarity 100.0%; Pred. No. 4.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVLAGK 7

Db 1665 gtvlagk 1671

RESULT 9

AAE00701 AAE00701 standard; Protein: 2123 AA.

AC AAE00701;

XX 02-JUL-2001 (first entry).

DE *Moraxella catarrhalis* outer membrane protein-106 (OMP106).

XX Outer membrane protein-106; OMP106; haemagglutination; vaccine;

KW bacterial infection; immunogen; cytotoxic; antitoxic;

KM passive immunisation.

XX *Moraxella catarrhalis*.

OS US6214981-B1.

PN 10-APR-2001.

PD 12-NOV-1997; 97US-0968685.

PF 03-MAY-1996; 96US-0642712.

PR (ANTE-) ANTEX BIOLOGICS INC.

XX (ANTE-) ANTEX BIOLOGICS INC.

PI Tucker K, Plosila L, Tillman DF;

XX WPI: 2001-281002/29.

DR N-PSDB; AAD04029.

PT Novel nucleotide sequences encoding *Moraxella catarrhalis* outer

PT membrane protein-106 polypeptide, useful for diagnosis of bacterial

PT infections and as vaccine against *Moraxella catarrhalis* infection of

PT mammals

PS Claim 7; Column 53-64; 49pp; English.

XX The present sequence is haemagglutinating *Moraxella catarrhalis* outer

CC membrane protein-106 (OMP106). The OMP106 is used as a therapeutic

CC and prophylactic vaccine against *M. catarrhalis* infections of mammals.

CC It is used for diagnosis of bacterial infections and as reagents for

CC clinical or medical diagnosis of *M. catarrhalis* infections and for

CC scientific research on the properties of pathogenicity, virulence and

CC infectivity of *M. catarrhalis*. It is also used as a probe to identify

CC the presence of *M. catarrhalis* in biological specimens and to identify

CC other bacteria that encode a polypeptide related to *M. catarrhalis*

CC OMP106. OMP106-derived polypeptides are used as ligands to detect

CC antibodies elicited in response to *M. catarrhalis* infections and also

CC as immunogens for inducing *M. catarrhalis*-specific antibodies which are

CC useful in immunoassays to detect *M. catarrhalis* in biological specimens.

CC Cytotoxic antibodies are useful in passive immunisations against

CC *M. catarrhalis*.

SO Sequence 2123 AA:

Query Match 87.8%; Score 36; DB 22; Length 2123;

Best Local Similarity 100.0%; Pred. No. 4.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVLAGK 7

Db 1735 gtvlagk 1741

RESULT 10

AAE69136 AAE69136 standard; Protein: 2314 AA.

AC AAE69136;

XX 24-APR-2001 (first entry)

DE *M. catarrhalis* Iesi 200kDa protein SEQ ID NO:11.

XX *Moraxella catarrhalis* strain Q8; major outer membrane protein;

KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

KM otitis media; detection.

XX *Moraxella catarrhalis*.

OS WO200107619-A1.

PN 01-FEB-2001.

PD 26-JUL-2000; 2000WC-CA00870.

PF 27-JUL-1999; 99US-0361619.

PR (CONN-) CONNAUGHT LAB LTD.

XX Loomore SM, Sasaki K, Yang Y, Klein MH;

PI WPI: 2001-159722/16.

DR N-PSDB; AAF59105.

XX New nucleic acid encoding *Moraxella catarrhalis* outer membrane protein,

PT useful in protective vaccines and for diagnosis

PT Claim 1; Fig 5A-Y; 247bp; English.

XX The present invention describes an isolated and purified nucleic acid (I)

CC that encodes a 200 kDa outer membrane protein of *Moraxella catarrhalis*.

CC The 200 kDa outer membrane protein (II) has antibacterial activity and

CC can be used in vaccines. (II), and its truncated versions, are used as

CC immunogenic compositions and vaccines to protect against *M. catarrhalis*

CC infections, particularly otitis media in humans. (II) is also used as

CC antigen in immunoassays for detecting specific antibodies (Ab), and to

CC generate Ab. (I) are used for recombinant production of (II) and its

CC fragments are used as probes for identifying/cloning 200 kDa protein

CC genes from other strains, and for diagnostic detection of *M. catarrhalis*.

CC (I) makes possible production of large amount of recombinant immunogens.

CC Expression of truncated versions of (II) reduces toxicity of the protein

CC towards the Escherichia coli host. The present sequence represents the
CC M. catarrhalis lesl 200kDa protein, which is given in the exemplification
CC of the present invention.

XX Sequence 2314 AA;

Query Match 87.8%; Score 36; DB 22; Length 2314;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVLGSK 7
1111111
Db 1926 gtvl9gk 1932

RESULT 11

AAW18786
ID AAW18786 standard; Protein; 509 AA.

XX AAW18786;

XX 10-DEC-1997 (first entry)

DE Neisseria adhesion protein (from OrfB).

XX Neisseria gonorrhoeae; adhesion; lipoprotein; OrfA; OrfI; OrfB.

OS Neisseria gonorrhoeae.

XX DE19534579-A1.

XX 20-MAR-1997.

XX 18-SEP-1995; 95DE-1034579.

XX 18-SEP-1995; 95DE-1034579.

PA (PLAC) MAX PLANCK GRS FOERDERUNG WISSENSCHAFTEN.

PI Fischer E, Maier J, Meyer TF, Rudel T, Scheuerpflug I;

DR WPL; 1997-180942/17.

PT Nucleic acids encoding Neisseria adhesion proteins - for therapeutic
PS and diagnostic use

XX Claim 40; Page 11-13; 20pp; German.

CC OrfA and OrfB in complexes with the protein PilC are capable of
CC adhering to human cells.

CC Products obtained from the DNA are useful in medicaments,
CC diagnostic compns. and vaccines, esp. for treatment of

CC Neisseria gonorrhoea and N. meningitidis infections.

XX Sequence 509 AA;

Query Match 85.4%; Score 35; DB 18; Length 509;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGSK 8
1111111
Db 326 gtvl9gk 333

RESULT 12

ABBI7879
ID ABBI7879 standard; Protein; 109 AA.

XX AC ABBI7879;

DT 23-JAN-2002 (first entry)
XX Human nervous system related polypeptide SEQ ID NO 6536.
DE Human
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antischizoid; antianemic; antithrombotic; cancer;
KW antipneumatic; hepatotropic; cerebroprotective; anti-inflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine.
XX Homo sapiens.
XX WO200159063-A2.
XX 16-AUG-2001.
XX 17-JAN-2001; 2001WO-US01334.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 03-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 20-OCT-2000; 2000US-0242221.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI, 2001-541565/60.
 DR N-PSDB; ABA14205.
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX
 PS Claim 11: SEQ ID NO 6536; 1701pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
 CC (ABA14678-ABA18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 109 AA:
 OY 1 GTVVGKK 8
 Db 5 gsvlggqk 12
 RESULT 13:
 AAG61517
 ID AAG61517 standard; Protein: 184 AA.
 XX
 AC AAG61517;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 79798.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 OS
 XX
 FN EPI033405-A2.
 XX

Query Match 80.5%; Score 33; DB 22; Length 109;
 Best Local Similarity 75.0%; Pred. No. 68;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130444.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139465.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145215.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.

PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 23-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match 80.58; Score 33; DB 21; Length 184;
Best Local Similarity 87.58; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGKK 8
Db 25 gtlvlgkk 32

RESULT 14

AG57078
ID AG57078 standard; Protein: 185 AA.

XX AC AG57078;

DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 73506.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 9905-0121825.

PR 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.

PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 23-APR-1999; 9905-0130891.
PR 28-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 04-MAY-1999; 9905-0132407.
PR 05-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132863.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0136229.
PR 25-MAY-1999; 9905-0136621.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137228.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.

PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159884.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

Best local Similarity 87.5%; Score 33; DB 21; Length 185;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGKK 8
Db 25 GTVLGKK 32

RESULT 15

AAG61516 standard; Protein; 190 AA.

XX AAG61516;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 79797.

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132464.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138547.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139465.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 26-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159338.
PR 18-OCT-1999; 99US-0159384.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 80.5%; Score 33; DB 21; Length 190;

Best Local Similarity 87.5%; Pred. NO. 1.2e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGCK 8
DB 31 gtlgqkk 38

Search completed: July 30, 2002, 15:34:12
Job time: 407 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:34:51 ; Search time 33.83 seconds
(without alignments)
5.776 Million cell updates/sec

Title: US-09-813-214A-2

Perfect score: 41

Sequence: 1 GTVLGKK 8

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 41 | 100.0 | 8 | 4 | US-08-968-685A-2 |
| 2 | 36 | 87.8 | 7 | 4 | US-08-968-685A-11 |
| 3 | 36 | 87.8 | 2048 | 4 | US-09-268-347-48 |
| 4 | 36 | 87.8 | 2123 | 4 | US-08-968-685A-10 |
| 5 | 36 | 87.8 | 2314 | 4 | US-08-268-347-49 |
| 6 | 32 | 78.0 | 174 | 4 | US-08-686-968C-10 |
| 7 | 32 | 78.0 | 363 | 6 | 5223606-6 |
| 8 | 32 | 78.0 | 933 | 4 | US-08-764-870-14 |
| 9 | 32 | 78.0 | 933 | 4 | US-08-980-115-14 |
| 10 | 31 | 75.6 | 134 | 4 | US-09-367-953B-120 |
| 11 | 31 | 75.6 | 1144 | 3 | US-08-726-214-6 |
| 12 | 30 | 73.2 | 20 | 2 | US-08-909-433A-11 |
| 13 | 30 | 73.2 | 174 | 5 | PCT-US95-10904-51 |
| 14 | 30 | 73.2 | 282 | 4 | US-09-120-365-79 |
| 15 | 30 | 73.2 | 282 | 4 | US-09-515-039-79 |
| 16 | 30 | 73.2 | 339 | 1 | US-08-208-007A-11 |
| 17 | 30 | 73.2 | 339 | 1 | US-08-330-121B-7 |
| 18 | 30 | 73.2 | 339 | 4 | US-08-860-255A-7 |
| 19 | 30 | 73.2 | 339 | 5 | PCT-US95-13820-7 |
| 20 | 30 | 73.2 | 360 | 1 | US-08-361-708-4 |
| 21 | 30 | 73.2 | 360 | 1 | US-08-536-277-4 |
| 22 | 30 | 73.2 | 361 | 1 | US-08-361-708-3 |
| 23 | 30 | 73.2 | 361 | 1 | US-08-314-309A-25 |
| 24 | 30 | 73.2 | 361 | 1 | US-08-536-277-3 |
| 25 | 30 | 73.2 | 822 | 2 | US-08-474-067-9 |
| 26 | 30 | 73.2 | 822 | 2 | US-08-474-068A-9 |
| 27 | 30 | 73.2 | 822 | 2 | US-08-472-481-8 |

| | | | | | | |
|----|----|------|------|---|-------------------|-------------------|
| 28 | 30 | 73.2 | 1313 | 2 | US-08-244-537-2 | Sequence 2, Appl |
| 29 | 30 | 73.2 | 2325 | 3 | US-08-417-089-6 | Sequence 6, Appl |
| 30 | 30 | 73.2 | 2325 | 4 | US-08-695-651-6 | Sequence 6, Appl |
| 31 | 30 | 73.2 | 2325 | 4 | US-08-930-285-6 | Sequence 6, Appl |
| 32 | 30 | 73.2 | 2325 | 4 | US-08-695-421-6 | Sequence 6, Appl |
| 33 | 30 | 70.7 | 19 | 4 | US-09-025-596-8 | Sequence 8, Appl |
| 34 | 30 | 70.7 | 72 | 1 | US-08-253-155A-48 | Sequence 48, Appl |
| 35 | 30 | 70.7 | 108 | 4 | US-09-240-274-51 | Sequence 51, Appl |
| 36 | 30 | 70.7 | 108 | 4 | US-09-240-274-52 | Sequence 52, Appl |
| 37 | 30 | 70.7 | 108 | 4 | US-09-240-274-53 | Sequence 53, Appl |
| 38 | 30 | 70.7 | 108 | 4 | US-09-240-274-166 | Sequence 166, App |
| 39 | 30 | 70.7 | 1276 | 1 | US-08-222-616-24 | Sequence 24, Appl |
| 40 | 30 | 70.7 | 1276 | 5 | PCT-US95-04228-24 | Sequence 24, Appl |
| 41 | 29 | 70.7 | 1580 | 2 | US-08-804-227C-11 | Sequence 11, Appl |
| 42 | 29 | 70.7 | 1580 | 2 | US-08-804-198-5 | Sequence 5, Appl |
| 43 | 29 | 70.7 | 1891 | 2 | US-08-804-227C-12 | Sequence 12, Appl |
| 44 | 29 | 70.7 | 1891 | 2 | US-08-804-198-6 | Sequence 6, Appl |
| 45 | 28 | 68.3 | 163 | 1 | US-08-044-618-6 | Sequence 6, Appl |

ALIGNMENTS

RESULT 1
US-08-968-685A-2
Sequence 2, Application US/08968685A
Patent No. 6214981

GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH

APPLICANT: PLOSILA, LAURA

TITLE OF INVENTION: MORAXELLA CAVARRHALIS OUTER MEMBRANE
PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/968,685A

FILING DATE: No. 6214981ember 12, 1997

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Baldwin, Geraldine F.

REGISTRATION NUMBER: 31,232

REFERENCE/DOCKET NUMBER: 7969-060

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ. ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 8 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: peptide

FRAGMENT TYPE: internal

US-08-968-685A-2

Query Match 100.0%; Score 41; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches: 8; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
Gy 1 GTVLGKK 8

Db 1 GTVLGSK 8

RESULT 2
US-08-968-685A-11
Sequence 11, Application US/08968685A
Patent No. 6214981

GENERAL INFORMATION:
APPLICANT: TUCKER, KENNETH
APPLICANT: PLOSTIA, LAURA
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,685A
FILING DATE: No. 6214981ember 12, 1997

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal

US-08-968-685A-11

Query Match 87.8%; Score 36; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGSK 7
Db 1 GTVLGSK 7

RESULT 3
US-09-268-347-48
Sequence 48, Application US/09268347
Patent No. 6335182
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
FILE REFERENCE: 1038-860
CURRENT APPLICATION NUMBER: US/09/268,347
CURRENT FILING DATE: 1999-03-16
NUMBER OF SEQ ID NOS: 54
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 48
LENGTH: 2048
TYPE: PRT

ORGANISM: Haemophilus influenzae
US-09-268-347-48

Query Match 87.8%; Score 36; DB 4; Length 2048;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGSK 7
Db 1660 GTVLGSK 1666

RESULT 4
US-08-968-685A-10
Sequence 10, Application US/08968685A
Patent No. 6214981
GENERAL INFORMATION:
APPLICANT: TUCKER, KENNETH
APPLICANT: PLOSTIA, LAURA
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,685A
FILING DATE: No. 6214981ember 12, 1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2123 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide

US-08-968-685A-10

Query Match 87.8%; Score 36; DB 4; Length 2123;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGSK 7
Db 1735 GTVLGSK 1741

RESULT 5
US-09-268-347-49
Sequence 49, Application US/09268347
Patent No. 6335182
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS

FILE REFERENCE: 1038-860
CURRENT APPLICATION NUMBER: US/09/268,347
CURRENT FILING DATE: 1999-03-16
NUMBER OF SEQ ID NOS: 54
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 49
LENGTH: 2314
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-268-347-49

Query Match 87.8%; Score 36; DB 4; Length 2314;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGK 7
|||||
DB 1926 GTVLGK 1932

RESULT 6
US-08-686-968C-10
Sequence 10, Application US/08686968C
Patent No. 6221361
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
APPLICANT: Junker, David E.
TITLE OF INVENTION: Recombinant Swinepox Virus
FILE REFERENCE: 39119-H/JML
CURRENT APPLICATION NUMBER: US/08/686,968C
CURRENT FILING DATE: 1996-07-25
NUMBER OF SEQ ID NOS: 231
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 174
TYPE: PRT
ORGANISM: Porcine reproductive and respiratory syndrome virus
US-08-686-968C-10

Query Match 78.0%; Score 32; DB 4; Length 174;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGK 8
|||||
DB 153 GTVLGK 160

RESULT 7
5223606-6
Patent No. 5223606
APPLICANT: BLAUDIN DE THE, HUGHES; MARCHIO, AGNES; TIOLLAIS,
PIERRE-DEJAN, ANNE
TITLE OF INVENTION: STEROID/THYROID HORMONE RECEPTOR-RELATED
PROTEIN INAPPROPRIATELY EXPRESSED IN HUMAN HEPATOCELLULAR CARCINOMA
NUMBER OF SEQUENCES: 11
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/134,130
FILING DATE: 17-DEC-1987
PRIOR APPLICATION DATA:
SEQ ID NO: 6;
LENGTH: 363
5223606-6

Query Match 78.0%; Score 32; DB 6; Length 363;
Best Local Similarity 75.0%; Pred. No. 89;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGK 8
|||||

DB 65 GMYLGK 72

RESULT 8
US-08-764-870-14
Sequence 14, Application US/08764870
Patent No. 6236946
GENERAL INFORMATION:
APPLICANT: Scanlan, Thomas S
APPLICANT: Baxter, John D
APPLICANT: Fletchick, Robert J
APPLICANT: Wagner, Richard L
APPLICANT: Kushner, Peter J
APPLICANT: Apriletti, James W
APPLICANT: West, Brian
TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooley Godward
STREET: Five Palo Alto Square, 3000 El Camino Real
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,870
FILING DATE: 13-DEC-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,540
FILING DATE: 13-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,543
FILING DATE: 13-DEC-1995
APPLICATION DATA:
PRIOR APPLICATION NUMBER: US 60/008,606
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Nakamura, Jackie N
REGISTRATION NUMBER: 35,966
REFERENCE/DOCKET NUMBER: UCAL-246/01US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)843-5000
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 933 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-764-870-14

Query Match 78.0%; Score 32; DB 4; Length 933;
Best Local Similarity 75.0%; Pred. No. 2,4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGK 8
|||||
DB 631 GMYLGK 638

RESULT 9
US-08-980-115-14
Sequence 14, Application US/08980115
Patent No. 626622
GENERAL INFORMATION:

APPLICANT: Scanlan, Thomas S.
APPLICANT: Baxter, John D.
APPLICANT: Fletchick, Robert J.
APPLICANT: Wagner, Richard L.
APPLICANT: Kushner, Peter J.
APPLICANT: Apiletti, James W.
APPLICANT: West, Brian L.
APPLICANT: Shlau, Andrew K.
TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS
FILE REFERENCE: UCAL-246/0205
CURRENT APPLICATION NUMBER: US/08/980,115
CURRENT FILING DATE: 1997-11-26
EARLIER APPLICATION NUMBER: 08/764,870
EARLIER FILING DATE: 1996-12-13
EARLIER APPLICATION NUMBER: 60/008,606
EARLIER FILING DATE: 1995-12-14
EARLIER APPLICATION NUMBER: 60/008,543
EARLIER FILING DATE: 1995-12-13
EARLIER APPLICATION NUMBER: 60/008,540
EARLIER FILING DATE: 1995-12-13
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 14
LENGTH: 933
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: DOMAIN
LOCATION: (659)..(918)
OTHER INFORMATION: minimal ligand binding domain
US-08-980-115-14

Query Match 78.0%; Score 32; DB 4; Length 933;
Best Local Similarity 75.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGSK 8
DB 631 GTVLGSK 638

RESULT 10
US-09-367-953B-120
Sequence 120, Application US/09367953B
Patent No. 6287572
GENERAL INFORMATION:
APPLICANT: Kingsman, Alan J
APPLICANT: Kingsman, Susan M
TITLE OF INVENTION: Anti-HIV peptides and proteins
FILE REFERENCE: 550-154
CURRENT APPLICATION NUMBER: US/09/367,953B
CURRENT FILING DATE: 1999-08-24
PRIOR APPLICATION NUMBER: PCT/GB98/00563
PRIOR FILING DATE: 1998-02-23
PRIOR APPLICATION NUMBER: GB 9703802.0
PRIOR FILING DATE: 1997-02-24
NUMBER OF SEQ ID NOS: 131
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 120
LENGTH: 134
TYPE: PRT
ORGANISM: Human immunodeficiency virus type 2
US-09-367-953B-120

Query Match 75.6%; Score 31; DB 4; Length 134;
Best Local Similarity 75.0%; Pred. No. 49;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGSK 8
DB 5 GSVLGSK 12

RESULT 11
US-08-726-214-6
Sequence 6, Application US/08726214
Patent No. 6107076
GENERAL INFORMATION:
APPLICANT: Tang, Wei-Jen
APPLICANT: Gilman, Alfred G.
TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE
TITLE OF INVENTION: AND USES THEREFOR
NUMBER OF SEQUENCE: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,214
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/005,498
FILING DATE: 04-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: UTSD:450
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1144 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-726-214-6

Query Match 75.6%; Score 31; DB 3; Length 1144;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVLG 6
DB 418 GTVLG 423

RESULT 12
US-08-909-433A-1
Sequence 1, Application US/08909433A
Patent No. 5981698
GENERAL INFORMATION:
APPLICANT: Andrew Bittalin
TITLE OF INVENTION: ACTIVE FRAGMENT OF
TITLE OF INVENTION: MODIFIED INSECTICIDAL CRYSTAL PROTEINS OF
TITLE OF INVENTION: BACILLUS THURINGIENSIS-RELATED POLY-
TITLE OF INVENTION: PEPTIDES FOR ANTIMICROBIAL AND STERILIZING
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bryan Cave LLP
STREET: 245 Park Avenue
CITY: New York
STATE: New York

COUNTRY: USA
ZIP: 10167-0034
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50, 800kb storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Microsoft Word 7.0.
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/909,433A
FILING DATE: 03 JUN 1997
ATTORNEY/AGENT INFORMATION:
NAME: Leo G. Lenna
REGISTRATION NUMBER: 42,796
REFERENCE/DOCKET NUMBER: C36599/105967
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 692-1839
TELEFAX: (212) 692-1900
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 Amino Acids
TYPE: Amino Acid
TOPOLOGY: UNKNOWN
MOLECULE TYPE: Peptide
US-08-909-433A-1

Query Match 73.2% Score 30; DB 2; Length 20;
Best Local Similarity 75.0%; Pred. No. 11;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGKK 8
11111111
DB 3 GTVLGKK 10

RESULT 13
PCT-US95-10904-51
Sequence 51, Application PC/TUS9510904
GENERAL INFORMATION:
APPLICANT: PAUL, PREM S.
APPLICANT: MENG, XIANG-JIN
APPLICANT: HALBUR, PATRICK G.
APPLICANT: MOROZOV, IGOR
APPLICANT: LOM, MELISSA A.
TITLE OF INVENTION: A POLYNUCLEIC ACID ISOLATED FROM A
TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV),
TITLE OF INVENTION: A PROTEIN ENCODED BY THE POLYNUCLEIC ACID, A VACCINE
TITLE OF INVENTION: PREPARED FROM OR CONTAINING THE POLYNUCLEIC ACID OR
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10904
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/131,625
FILING DATE: 05-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M.P.
REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 4625-021-55X CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 174 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-10904-51

Query Match 73.2% Score 30; DB 5; Length 174;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 VLGGKK 8
11111111
DB 155 VLGGKK 160

RESULT 14
US-09-120-365-79
Sequence 79, Application US/09120365
GENERAL INFORMATION:
PATENT NO. 610314
APPLICANT: Natori, Shunji
TITLE OF INVENTION: NEW PROTEASE
FILE REFERENCE: 32290-144749
CURRENT APPLICATION NUMBER: US/09/120,365
CURRENT FILING DATE: 1998-07-22
EARLIER APPLICATION NUMBER: JP 9-333 474
EARLIER FILING DATE: 1997-11-18
NUMBER OF SEQ ID NOS: 101
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 79
LENGTH: 282
TYPE: PRT
ORGANISM: Homo sapien cathepsin B
US-09-120-365-79

Query Match 73.2% Score 30; DB 3; Length 282;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTVLGKK 8
11111111
DB 60 GTVLGKK 67

RESULT 15
US-09-515-039-79
Sequence 79, Application US/09515039
PATENT NO. 621459
GENERAL INFORMATION:
APPLICANT: Natori, Shunji
TITLE OF INVENTION: NEW PROTEASE
FILE REFERENCE: 32290-144749
CURRENT APPLICATION NUMBER: US/09/515,039
CURRENT FILING DATE: 2000-03-06
EARLIER APPLICATION NUMBER: JP 9-333 474
EARLIER FILING DATE: 1997-11-18
NUMBER OF SEQ ID NOS: 101
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 79
LENGTH: 282
TYPE: PRT
ORGANISM: Homo sapien cathepsin B
US-09-515-039-79

Query Match 73.2%; Score 30; DB 4; Length 282;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTVLCGKK 8
|||
Db 60 GTFLGPK 67

Search completed: July 30, 2002, 15:34:52
Job time: 237 sec

A;Accession: A49826
A;Molecule type: mRNA

A;ResIdues: 1-339 <QI2>

A:Cross-references: GB:S69034; NID:g239906; PIDN:AAB20536.1; PID:g239907
 A:Experimental source: B16a melanoma
 A:Note: Sequence extracted from B16a melanoma (NCIN:69034, NCHP:69039)
 A:Note: three mRNA forms from B16a melanoma cells were found to have identical coding se
 Proc. Natl. Acad. Sci. U.S.A. 83, 7721-7725, 1986
 R:Chan, S.J.; San Segundo, B.; McCormick, M.B.; Steiner, D.F.
 A:Title: Nucleotide and predicted amino acid sequences of cloned human and mouse preproc
 A:Reference number: A26498; MUID:87017021
 A:Accession: B26498
 A:Molecule type: mRNA
 A:Residues: 1-159, 'N', 161-173, 'D', 175-176, 'I', 178-283, 'V', 285-339 <CHN>
 A:Cross-references: GB:M14222; NID:g192841; PIDN:AAA37494.1; PID:g309202
 R:Ferrera, M.; Wojcik, F.; Rhaisst, H.; Mordier, S.; Roux, M.P.; Bechet, D.
 FEBS Lett. 273, 195-199, 1990
 A:Title: Gene structure of mouse cathepsin B.
 A:Reference number: S12901; MUID:91032179
 A:Accession: S12901
 A:Molecule type: DNA
 A:Residues: 1-159, 'N', 161-173, 'D', 175-176, 'I', 178-283, 'V', 285-339 <FER>
 R:Fluimert, C.; Closs, E.I.; Silbermann, M.; Erfle, V.; Strauss, P.G.
 Gene 103, 259-261, 1991
 A:Title: Isolation of a cathepsin B-encoding cDNA from murine osteogenic cells.
 A:Reference number: PS0360; MUID:91365255
 A:Accession: PS0360
 A:Status: translation not shown
 A:Molecule type: mRNA
 A:Residues: 314-339 <FRI>
 A:Cross-references: EMBL:X54966; NID:g50596; PIDN:CAA38713.1; PID:g50597
 C:Genetics:
 A:Note: single copy gene
 C:Function:
 A:Description: broad specificity endopeptidase and peptidyl dipeptidase
 A:Pathway: intracellular protein degradation
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; glycoprotein; hydrolase; lysosome; protein degradation
 F:1-17/Domain: signal sequence #status predicted <SIG>
 F:80-339/Product: cathepsin B #status predicted <MAT>
 F:8, 192/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:93-122, 105-150, 141-207, 142-146, 179-211, 187-198/Disulfide bonds: #status predicted
 F:108, 278/Active site: Cys, His, Asn #status predicted

Query Match 85.4%; Score 35; DB 1; Length 339;
 Best Local Similarity 87.5%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVVGCK 8
 |||||
 Db 60 GTVVGCK 67

RESULT 3
 S63533
 C:Species: Entamoeba histolytica
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
 C:Accession: S63533; S63474
 R:Binder, M.; Ortner, S.; Erben, H.; Scheiner, O.; Wiedermann, G.; Valenta, R.; Duchene,
 Eur. J. Biochem. 233, 976-981, 1995
 A:Title: The basic isoform of profilin in pathogenic Entamoeba histolytica: cDNA cloning
 A:Reference number: S63474; MUID:96085166
 A:Accession: S63474
 A:Molecule type: mRNA
 A:Residues: 1-130 <BIN>
 A:Cross-references: EMBL:X90911; NID:g1070154; PID:g1070155
 A:Accession: S63474
 A:Molecule type: protein
 A:Residues: 74-84; 86-94 <BIN>
 C:Superfamily: profilin
 C:Keywords: actin binding

Query Match 82.9%; Score 34; DB 2; Length 130;

Best Local Similarity 75.0%; Pred. No. 24;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVVGCK 8
 |||||
 Db 61 GTVVGCK 68

RESULT 4

H82745
 acyl-[ACP]-UDP-N-acetylglucosamine XF0918 [Imported] - Xylella fastidiosa (strain 9a5
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: H82745
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
 Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: H82745

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-214 <SIM>

A:Cross-references: GB:AE003931; GB:AE003849; NID:g9105834; PIDN:AAF83728.1; GSPDB:GN

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carier
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facinca, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
 chado, M.A.; Medeira, A.M.B.N.; Medeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, C.
 Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
 M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF0918

Query Match 82.9%; Score 34; DB 2; Length 214;
 Best Local Similarity 85.7%; Pred. No. 39;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVVGCK 7
 |||||
 Db 59 GTVVGCK 65

RESULT 5

B97860
 hypothetical protein RC1282 [Imported] - Rickettsia conorii (strain Malish 7)

C:Species: Rickettsia conorii

C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001

C:Accession: B97860
 R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barde, V.; Samson, D.;

Science 293, 2093-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.

A:Reference number: A97700; MUID:21442074; PMID:11557893

A:Accession: B97860

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-227 <KUR>

A:Cross-references: GB:AE006914; PIDN:ALU03820.1; PID:g15620420; GSPDB:GN00173

C:Genetics:

C:Superfamily: Rickettsia prowazekii hypothetical protein RP828

Query Match 80.5%; Score 33; DB 2; Length 227;

Best Local Similarity 71.4%; Pred. No. 64;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGSK 7
||:||||
Db 82 GTVLGSK 88

RESULT 6

C95846

hypothetical protein [imported] - *Sinorhizobium meliloti* (strain 1021) megaplasmid pSymb

C:Species: *Sinorhizobium meliloti*

C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: C95846

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernat

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,683-Kb pSymb megaplasmid from the N2-fixing endo

A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: C95846

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-244 <KUR>

A:Cross-References: GB:AL591985; PIDN:CA048435.1; PID:915139907; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid pSymb

R:Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Laure,

heault, P.; Vandebol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: SMD20032

A:Genome: plasmid

Query Match 80.5%; Score 33; DB 2; Length 244;
Best Local Similarity 85.7%; Pred. No. 69;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGSK 7
||:||||
Db 214 GTVLGSK 220

RESULT 7

S63385

hypothetical protein YNR054c - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein N3491

C:Species: *Saccharomyces cerevisiae*

C>Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 05-Nov-1999

C:Accession: S63385

R:Pohl, T.M.

submitted to the Protein Sequence Database, April 1996

A:Reference number: S63346

A:Accession: S63385

A:Molecule type: DNA

A:Residues: 1-316 <POH>

A:Cross-References: EMBL:Z71669; NID:g1302571; PID:e239595; PID:g1302572; GSPDB:GN00014;

A:Experimental source: strain S288C

C:Genetics:

A:Gene: MIPS:YNR054c

A:Map position: 14R

Query Match 80.5%; Score 33; DB 2; Length 316;
Best Local Similarity 62.5%; Pred. No. 88;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGSK 8
||:||||

Db 189 GNIGGKK 196

RESULT 8

S31735

NAD+ ADP-ribosyltransferase (EC 2.4.2.30) - African clawed frog (fragment)

N:Alternate names: poly ADP-ribose polymerase

C:Species: *Xenopus laevis* (African clawed frog)

C>Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000

C:Accession: S31735; P0495

R:Sautier-Le Drian, B.M.

submitted to the EMBL Data Library, May 1992

A:Reference number: S31735

A:Accession: S31735

A:Molecule type: mRNA

A:Residues: 1-998 <SAD>

A:Cross-References: EMBL:Z12139; NID:964967; PIDN:CA078126.1; PID:91334661

R:Ozawa, Y.; Uchida, K.; Uchida, M.; Aml, Y.; Kushida, S.; Okada, N.; Miwa, M.

Biochem. Biophys. Res. Commun. 193, 119-125, 1993

A:Title: Isolation of cDNAs encoding the catalytic domain of poly(ADP-ribose)polymera

A:Reference number: P04954; MUID:93277538

A:Accession: P0495

A:Molecule type: mRNA

A:Residues: 742-745, 'E', 747-876 <OZA>

C:Comment: This zinc-finger protein plays a role in DNA repair, cell growth, and diff

C:Superfamily: NAD+ ADP-ribosyltransferase

C:Keywords: DNA binding; glycosyltransferase; hexosyltransferase; NAD; nucleus; pento

Query Match 80.5%; Score 33; DB 2; Length 998;
Best Local Similarity 75.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGSK 8
||:||||
Db 576 GTVLGSK 583

RESULT 9

T38477

very hypothetical protein SPAC29A4.12c - fission yeast (*Schizosaccharomyces pombe*)

C:Species: *Schizosaccharomyces pombe*

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T38477

R:Brown, D.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, June 1997

A:Reference number: Z21796

A:Accession: T38477

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-150 <BRG>

A:Cross-References: EMBL:Z97210; PIDN:CA010139.1; GSPDB:GN000066; SPDB:SPAC29A4.12c

A:Experimental source: strain 972h-; cosmid c29A4

C:Genetics:

A:Gene: SPDB:SPAC29A4.12c

A:Map position: 1

Query Match 78.0%; Score 32; DB 2; Length 150;
Best Local Similarity 62.5%; Pred. No. 68;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGSK 8
||:||||
Db 101 GSIFGSK 108

RESULT 10

D64694

hypothetical protein HP1396 - *Helicobacter pylori* (strain 26695)

C:Species: *Helicobacter pylori*

C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999

C:Accession: D64694

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McKenna, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Wathey, L. Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser, C. A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*. A:Reference number: A64520; MUID:97394467

A:Accession: D64694

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-288 <TOM>

A:Cross-references: GB:AE000639; GB:AE000511; NID:g2314560; PIDN:AA08441.1; PID:g231457

Query Match
Best Local Similarity 78.0%; Score 32; DB 2; Length 288;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TVLGGK 8
Db 219 TVLGGK 225

RESULT 11
A96953
methyl-accepting chemotaxis protein [imported] - *Clostridium acetobutylicum*
C:Species: *Clostridium acetobutylicum*
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: A96953

R:Rolling, J.; Betton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum*
A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: A96953

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-337 <ROR>

A:Cross-references: GB:AE001437; PIDN:AAK78412.1; PID:g15023287; GSPDB:GN00168

A:Experimental source: *Clostridium acetobutylicum* ATCC824

C:Genetics:

A:Gene: CAC0432

Query Match
Best Local Similarity 78.0%; Score 32; DB 2; Length 337;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGK 7
Db 109 GTVLGK 115

RESULT 12
D75568
N-acetylaminic acid racemase - *Deinococcus radiodurans* (strain R1)
C:Species: *Deinococcus radiodurans*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: D75568

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: A75250; MUID:20036896

A:Accession: D75568

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-375 <WHI>

A:Cross-references: GB:AE001867; GB:AE000513; NID:g6457693; PIDN:AAF09631.1; PID:g645769

A:Experimental source: strain R1

C:Genetics:

A:Gene: D90044

A:Map position: 1
C:Superfamily: muconate cyclisomerase

Query Match
Best Local Similarity 78.0%; Score 32; DB 2; Length 375;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGK 8
Db 127 GTVLGK 134

RESULT 13
D96760
hypothetical protein T9L24.41 [imported] - *Arabidopsis thaliana*
C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D96760

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federapfel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, A.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzita, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A:Reference number: A6141; MUID:21016719

A:Accession: D96760

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-434 <SNO>

A:Cross-references: GB:AE005173; NID:g1120791; PIDN:AA030971.1; GSPDB:GN00141

C:Genetics:

A:Gene: T9L24.41

A:Map position: 1

Query Match
Best Local Similarity 78.0%; Score 32; DB 2; Length 434;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGK 8
Db 324 GTVLGK 331

RESULT 14
AB3107
conserved hypothetical protein PAA308 [imported] - *Pseudomonas aeruginosa* (strain PAO)
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: AB3107

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Adam, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laribig, K.; L.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen.
A:Reference number: A82950; MUID:20437337

A:Accession: AB3107

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-496 <SNO>

A:Cross-references: GB:AE004847; GB:AE004091; NID:g9950528; PIDN:AA07696.1; GSPDB:GN00141

A:Experimental source: strain PAO1

C:Genetics:

A:Gene: PAA308

Query Match
Best Local Similarity 78.0%; Score 32; DB 2; Length 496;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVLGKK 8

Db 487 GSLMGKK 494

RESULT 15

T06041

hypothetical protein T24A18.60 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999

C:Accession: T06041

R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voelt, M.; Robben, J.; Volckaert, G.; Be

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15484

A:Accession: T06041

A:Molecule type: DNA

A:Residues: 1-717 <BEV>

A:Cross-references: EMBL:AL035680; GSPDB:GN00062; ATSP:T24A18.60

A:Experimental source: cultivar Columbia; BAC clone T24A18

C:Genetics:

A:Gene: ATSP:T24A18.60

A:Map position: 4

A:Introns: 7/3; 239/1; 287/3; 398/3

Query Match

Best Local Similarity 78.0%; Score 32; DB 2; Length 717;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGKK 7

Db 207 GTVFGKK 213

Search completed: July 30, 2002, 15:35:52
Job time: 257 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:37:58 ; Search time 27.4 Seconds

(without alignments)
11.305 Million cell updates/sec

Title: US-09-813-214A-2

Perfect score: 41

Sequence: 1 GTVLGSK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 35 | 85.4 | 339 | 1 | CATB_MOUSE |
| 2 | 34 | 82.9 | 130 | 1 | PROF_ENTRI |
| 3 | 33 | 80.5 | 316 | 1 | Y8BV_YEAST |
| 4 | 33 | 80.5 | 998 | 1 | PPOL_XENLA |
| 5 | 32 | 78.0 | 150 | 1 | YDPC_SCHPO |
| 6 | 32 | 78.0 | 180 | 1 | PRGR_MACRO |
| 7 | 32 | 78.0 | 375 | 1 | RCOL1_ARATH |
| 8 | 32 | 78.0 | 377 | 1 | PRGR_SHEEP |
| 9 | 32 | 78.0 | 786 | 1 | PRGR_CHICK |
| 10 | 32 | 78.0 | 923 | 1 | PRGR_MOUSE |
| 11 | 32 | 78.0 | 923 | 1 | PRGR_MOUSE |
| 12 | 32 | 78.0 | 930 | 1 | PRGR_RAT |
| 13 | 32 | 78.0 | 933 | 1 | PRGR_HUMAN |
| 14 | 31 | 75.6 | 148 | 1 | YORX_PYRMO |
| 15 | 31 | 75.6 | 339 | 1 | CATB_RAT |
| 16 | 31 | 75.6 | 367 | 1 | YOV4_CAEEL |
| 17 | 31 | 75.6 | 372 | 1 | PEX3_MOUSE |
| 18 | 31 | 75.6 | 372 | 1 | PEX3_RAT |
| 19 | 31 | 75.6 | 373 | 1 | PEX3_CRIO |
| 20 | 31 | 75.6 | 373 | 1 | PEX3_HUMAN |
| 21 | 31 | 75.6 | 434 | 1 | HES1_YEAST |
| 22 | 31 | 75.6 | 475 | 1 | Y294_MYCPN |
| 23 | 31 | 75.6 | 521 | 1 | GAG_HYD2 |
| 24 | 31 | 75.6 | 559 | 1 | FRB7_WHEAT |
| 25 | 31 | 75.6 | 565 | 1 | DSBD_ECO57 |
| 26 | 31 | 75.6 | 565 | 1 | DSBD_ECOLI |
| 27 | 31 | 75.6 | 627 | 1 | GIDA_COXBU |
| 28 | 31 | 75.6 | 628 | 1 | GIDA_BUCAP |
| 29 | 31 | 75.6 | 631 | 1 | GIDA_BUCAP |
| 30 | 31 | 75.6 | 631 | 1 | GIDA_VIBCH |
| 31 | 31 | 75.6 | 635 | 1 | GIDA_SYNT3 |
| 32 | 31 | 75.6 | 1144 | 1 | CVA3_HUMAN |
| 33 | 31 | 75.6 | 1144 | 1 | CVA3_RAT |

J

| | | | | | |
|----|----|------|------|---|------------|
| 34 | 31 | 75.6 | 1914 | 1 | KMLS_HUMAN |
| 35 | 30 | 73.2 | 15 | 1 | ASPL_IACSN |
| 36 | 30 | 73.2 | 171 | 1 | VPX_IDV |
| 37 | 30 | 73.2 | 212 | 1 | REP2_KRUA |
| 38 | 30 | 73.2 | 269 | 1 | AROE_ARCFU |
| 39 | 30 | 73.2 | 299 | 1 | ISPA_ECOLI |
| 40 | 30 | 73.2 | 299 | 1 | YC21_METVA |
| 41 | 30 | 73.2 | 301 | 1 | CUT2_SCHPO |
| 42 | 30 | 73.2 | 325 | 1 | MOCG_RHIME |
| 43 | 30 | 73.2 | 328 | 1 | K6PF_MYCPN |
| 44 | 30 | 73.2 | 339 | 1 | CATB_HUMAN |
| 45 | 30 | 73.2 | 340 | 1 | CATB_CHICK |

ALIGNMENTS

| RESULT | ID | CATB_MOUSE | STANDARD | PRT | 339 AA |
|--------|---|-----------------------------------|----------|-----|--------|
| AC | P10605 | | | | |
| DT | 01-JUL-1989 | (Rel. 11, Created) | | | |
| DT | 01-FEB-1991 | (Rel. 17, Last sequence update) | | | |
| DT | 15-DEC-1998 | (Rel. 37, Last annotation update) | | | |
| DE | Cathepsin B precursor (EC 3.4.22.1) (Cathepsin B1). | | | | |
| GN | CNSB. | | | | |
| OS | Mus musculus (Mouse). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | | |
| OX | NCBI_Taxid=10090; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RX | MEDLINE=91190267; PubMed=2012677; | | | | |
| RA | Qian F., Franklater A., Chan S.J., Steiner D.F.; | | | | |
| RT | "The structure of the mouse cathepsin B gene and its putative promoter."; | | | | |
| RL | DNA Cell Biol. 10:159-168(1991). | | | | |
| RN | [2] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RX | MEDLINE=91032179; PubMed=2226854; | | | | |
| RA | Ferrara M., Wojcik F., Rhaissi H., Mordler S., Roux M.-P., Bechet D.; | | | | |
| RT | "Gene structure of mouse cathepsin B."; | | | | |
| RL | FEBS Lett. 273:195-199(1990). | | | | |
| RN | [3] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RX | MEDLINE=87017021; PubMed=346396; | | | | |
| RA | Chan S.J., San Segundo B., McCormick M.B., Steiner D.F.; | | | | |
| RT | "Nucleotide and predicted amino acid sequences of cloned human and mouse preprocathepsin B cDNAs."; | | | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 83:7721-7725(1986). | | | | |
| RN | [4] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RX | MEDLINE=92082172; PubMed=1746902; | | | | |
| RA | Qian F., Franklater A., Steiner D.F., Bajkowski A.S., Chan S.J.; | | | | |
| RT | "Characterization of multiple cathepsin B mRNAs in murine B16 melanoma."; | | | | |
| RL | Anticancer Res. 11:1445-1451(1991). | | | | |
| RN | [5] | | | | |
| RP | SEQUENCE OF 314-339 FROM N.A. | | | | |
| RX | MEDLINE=9136525; PubMed=1889751; | | | | |
| RA | Friemert C., Closs E.I., Silbermann M., Efile V., Strauss P.G.; | | | | |
| RT | "Isolation of a cathepsin B-encoding cDNA from murine osteogenic cells."; | | | | |
| RL | Gene 103:259-261(1991). | | | | |
| RN | [6] | | | | |
| RP | FUNCTION: THIOL PROTEASE WHICH IS BELIEVED TO PARTICIPATE IN INTRACELLULAR DEGRADATION AND TURNOVER OF PROTEINS. HAS ALSO BEEN IMPLICATED IN TUMOR INVASION AND METASTASIS. | | | | |
| CC | -1- CATALYTIC ACTIVITY: HYDROLASES PROTEINS, WITH A SPECIFICITY RESEMBLING THAT OF PAPAIN. | | | | |
| CC | -1- SUBUNIT: DIMER OF A HEAVY CHAIN AND A LIGHT CHAIN CROSS-LINKED BY A DISULFIDE BOND. | | | | |
| CC | -1- SUBCELLULAR LOCATION: Lysosomal. | | | | |
| CC | -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE | | | | |

| | |
|--------|-------------|
| Q15746 | homo sapien |
| P82648 | lactocacil1 |
| P24123 | lactocacil1 |
| P13779 | kluyveromy |
| O27857 | archaeog10b |
| P22939 | escherichia |
| Q58618 | methanococc |
| P21135 | schizosacch |
| P49304 | rhizobium m |
| P75476 | mycoplasma |
| P07858 | homo sapien |
| P43233 | gallus gall |

PAPAIN FAMILY OF THIOL PROTEASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 CC
 CC EMBL: M65210; AAA37375.1; -
 DR EMBL: M65263; AAA37375.1; JOINED.
 DR EMBL: M65264; AAA37375.1; JOINED.
 DR EMBL: M65265; AAA37375.1; JOINED.
 DR EMBL: M65266; AAA37375.1; JOINED.
 DR EMBL: M65267; AAA37375.1; JOINED.
 DR EMBL: M65268; AAA37375.1; JOINED.
 DR EMBL: M65269; AAA37375.1; JOINED.
 DR EMBL: M14222; AAA37494.1; -
 DR EMBL: X54966; CAA38713.1; -
 DR EMBL: S69034; AAB20536.1; -
 DR PIR: B26498; KMSB.
 DR PIR: S12901; S12901.
 DR PIR: A38458; A38458.
 DR PIR: A49826; A49826.
 DR HSP: P00787; 1THE.
 DR MEROPS: C01.060; -
 DR SWISS-2DPAGE: P10605; MOUSE.
 DR MGD: MGI:88561; Ct5b.
 DR InterPro: IPR000668; Peptidase.C1.
 DR InterPro: IPR00169; ThiolProt_act_site.
 DR Pfam: PF00112; Peptidase.C1; 1.
 DR PRINTS: PR00705; PAPAIN.
 DR PROSITE: PS00139; THOL_PROTEASE_CYS; 1.
 DR PROSITE: PS00639; THOL_PROTEASE_HIS; 1.
 DR PROSITE: PS00640; THOL_PROTEASE_ASN; 1.
 DR PROSITE: PS00640; THOL_PROTEASE_ASN; 1.
 DR Hydrolase: Thiol protease; Lysosome; Glycoprotein; Zymogen; Signal.
 KM
 FT SIGNAL 1 17
 FT PROPEP 18 79
 FT CHAIN 80 333
 FT CHAIN 80 333
 FT CHAIN 129 333
 FT CHAIN 129 333
 FT PROPEP 334 339
 FT ACT_SITE 108 108
 FT ACT_SITE 278 278
 FT ACT_SITE 298 298
 FT ACT_SITE 93 122
 FT DISULFID 105 150
 FT DISULFID 141 207
 FT DISULFID 142 146
 FT DISULFID 179 211
 FT DISULFID 187 198
 FT CARBOHYD 192 192
 FT CARBOHYD 160 160
 FT CONFLICT 174 174
 FT CONFLICT 177 177
 FT CONFLICT 284 284
 FT CONFLICT 284 284
 SQ SEQUENCE 339 AA; 37280 MW; 9F0A3CDF70A94040 CRC64;

Query Match 85.4%; Score 35; DB 1; Length 339;
 Best Local Similarity 87.5%; Pred. No. 13;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVIGGK 8
 DB 60 GTVIGGPK 67

RESULT 2
 PROF_ENTH1 STANDARD; PRT; 130 AA.
 AC P49230;

DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Profilin.
 OS Entamoeba histolytica.
 OC Eukaryota; Entamoebidae; Entamoeba.
 OX NCBI_Taxid=5759;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 74-84 AND 86-94.
 RC STRAIN=SFL-3;
 RX MEDLINE=96085166; PubMed=8521867;
 RA Balder M., Orthen S., Erden H., Scheiner O., Wiedermann G.,
 RA Valenta R., Duchene M.;
 RT "The basic isoform of profilin in pathogenic Entamoeba histolytica.
 RT cDNA cloning, heterologous expression, and actin-binding
 RT properties";
 RL Eur. J. Biochem. 233:976-981(1995)
 CC -1- FUNCTION: BINDS TO ACTIN AND AFFECTS THE STRUCTURE OF THE
 CC CYTOSKELETON. AT HIGH CONCENTRATIONS, PROFILIN PREVENTS THE
 CC POLYMERIZATION OF ACTIN, WHEREAS IT ENHANCES IT AT LOW
 CC CONCENTRATIONS. BY BINDING TO PIP2, IT INHIBITS THE FORMATION OF
 CC IP3 AND DG.
 CC -1- SUBUNIT: OCCURS IN MANY KINDS OF CELLS AS A COMPLEX WITH MONOMERIC
 CC ACTIN IN A 1:1 RATIO.
 CC -1- SIMILARITY: BELONGS TO THE PROFILIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 CC
 CC EMBL: X90911; CAA62418.1; -
 DR HSP: P19984; 2ACG.
 DR InterPro: IPR002097; Profilin.
 DR Pfam: PF00235; Profilin; 1.
 DR SMART: SM00392; PROF; 1.
 DR PROSITE: PS00414; PROFILIN; 1.
 DR Actin-binding: Cytoskeleton.
 KM Actin-binding: Cytoskeleton.
 SQ SEQUENCE 130 AA; 13237 MW; AF50A8E08FA00A6F CRC64;

Query Match 82.9%; Score 34; DB 1; Length 130;
 Best Local Similarity 75.0%; Pred. No. 8.6;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVIGGK 8
 DB 61 GTVIGGK 68

RESULT 3
 YN8V_YEAST STANDARD; PRT; 316 AA.
 ID YN8V_YEAST
 AC P53743;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical 36.4 kDa protein in POP2-HOL1 intergenic region.
 GN YN054C OR N3491.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_Taxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Pohl T.M.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: 271669; CAA96335.1; -
 DR SGD: S0005337; YNR054C.1
 DR InterPro: IPR003955; RRM_2.
 DR SMART: SM00362; RRM_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 316 AA; 36408 MW; 5F3B880CDEFC01F CRC64;

Query Match 80.5%; Score 33; DB 1; Length 316;
 Best Local Similarity 62.5%; Pred. No. 30;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGKK 8
 I::IIII
 Db 189 GNIGGKK 196

RESULT 4
 PPOL_XENLA STANDARD; PRT; 998 AA;
 AC P31669;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Poly (ADP-ribose) polymerase (EC 2.4.2.30) (PARP) (ADPRT) (NAD(+)-ADP-riboyltransferase) (Poly(ADP-ribose) synthetase) (Fragment).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Ovary;
 RA Saulier-Le Drian B.M.;
 RL Thesis (1992), University of Rennes, France.
 RT [2]
 RX MEDLINE-93277538; PubMed-8503897;
 RA Ozawa Y., Uchida K., Uchida M., Ami Y., Kushida S., Okada N., Miwa M.;
 RT "Isolation of cDNAs encoding the catalytic domain of poly(ADP-ribose) polymerase from xenopus laevis and cherry salmon using heterologous oligonucleotide consensus sequences";
 RL Blochem. Biophys. Res. Commun. 193:119-125(1993).
 CC -1- FUNCTION: POLY(ADP-RIBOSE) POLYMERASE MODIFIES VARIOUS NUCLEAR PROTEINS BY POLY(ADP-RIBOSYL)ATION. THE MODIFICATION IS DEPENDENT ON DNA AND IS INVOLVED IN THE REGULATION OF VARIOUS IMPORTANT CELLULAR PROCESSES SUCH AS DIFFERENTIATION, PROLIFERATION, AND TUMOR TRANSFORMATION AND ALSO IN THE REGULATION OF THE MOLECULAR EVENTS INVOLVED IN THE RECOVERY OF CELL FROM DNA DAMAGE.
 CC -1- CATALYTIC ACTIVITY: NAD(+) + [ADP-D-riboyl](N)-acceptor = nicotinamide + (ADP-D-riboyl)(N+1)-acceptor.
 CC -1- COFACTOR: ZINC, CONTAINS TWO MOLES OF ZINC PER MOLE OF PROTEIN.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN OVARY, OOCYTES, AND BRAIN. LOW IN LIVER.
 CC -1- MISCELLANEOUS: THE ADP-D-RIBOSYL GROUP OF NAD(+) IS TRANSFERRED TO AN ACCEPTOR CARBOXYL GROUP ON A HISTONE OR THE ENZYME ITSELF, AND FURTHER ADP-RIBOSYL GROUPS ARE TRANSFERRED TO THE 2'-POSITION OF THE TERMINAL ADENOSINE MOIETY, BUILDING UP A POLYMER WITH AN AVERAGE CHAIN LENGTH OF 20-30 UNITS.
 CC -1- SIMILARITY: BELONGS TO THE PARP FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 BCT DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: 212139; GAA78126.1; -
 DR EMBL: D13810; BAA02966.1; -
 DR PIR: S31735; S31735.
 DR HSSP: P26446; IAK6.
 DR InterPro: IPR001357; BRCT.
 DR InterPro: IPR001290; PARP.
 DR InterPro: IPR004102; PARP-reg.
 DR InterPro: IPR001510; Znf-PARP.
 DR Pfam: PF00553; BRC1.1.
 DR Pfam: PF00644; PARP.1.
 DR Pfam: PF02877; PARP-reg.1.
 DR Pfam: PF00645; zif-PARP.2.
 DR ProDom: PD004675; Znf-PARP.2.
 DR SMART: SM00292; BRC1.1.
 DR PROSITE: PS50172; BRC1.
 DR PROSITE: PS00347; PARP_ZN_FINGER_1; 2.
 DR PROSITE: PS50064; PARP_ZN_FINGER_2; 2.
 KW Transferase; Glycosyltransferase; NAD; DNA-binding; Nuclear protein; ADP-ribosylation; zinc-finger; zinc.
 FT NON_TER 1 1
 FT DNA_BIND <1 356
 FT DOMAIN 357 507
 FT DOMAIN 369 445
 FT DOMAIN 508 998
 FT ZN_FING 8 43
 FT ZN_FING 111 148
 FT DOMAIN 193 195
 FT DOMAIN 207 212
 FT DOMAIN 391 391
 FT MOD_RES 397 397
 FT MOD_RES 419 419
 FT MOD_RES 428 428
 FT MOD_RES 429 429
 FT MOD_RES 445 445
 FT MOD_RES 447 447
 FT MOD_RES 454 454
 FT MOD_RES 467 467
 FT MOD_RES 471 471
 FT MOD_RES 477 477
 FT MOD_RES 495 495
 FT MOD_RES 496 496
 FT MOD_RES 503 503
 FT CONFLICT 746 746
 SQ SEQUENCE 998 AA; 111126 MW; F5A25FA3366BAE7 CRC64;

Query Match 80.5%; Score 33; DB 1; Length 998;
 Best Local Similarity 75.0%; Pred. No. 85;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGKK 8
 I::IIII
 Db 576 GTVIGSKK 583

RESULT 5
 YDPC_SCHPO STANDARD; PRT; 150 AA;
 ID YDPC_SCHPO
 AC O14015;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Hypothetical 16.8 kDa protein C29A4.12c in chromosome I.
 GN SPAC29A4.12C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RL Brown D., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;
 Submitted (Jul-1997) to the EMBL/GenBank/DBJ databases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: 297210; CAB10139.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 150 AA; 16781 MW; 3CFPCDE23106361 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 150;
 Best Local Similarity 62.5%; Pred. No. 24;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGGRK 8
 I::IIII
 Db 101 GSIFGGRK 108

RESULT 6
 PROGR_MACEU STANDARD: PRT; 180 AA.
 AC P79373;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Progesterone receptor (PR) (Fragment).
 GN PGR OR NR3C3.
 OS Macropus eugenii (Tamar wallaby).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Metatheria; Dipriconodontia; Macropodidae; Macropus.
 OX NCBI_TaxID=9315;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Uterus;
 RX MEDLINE=96401260; PubMed=8807636;
 RA Lim-tio S.S., Keightley M.C., Fletcher T.P., Fuller P.J.;
 RT "The molecular basis of RU486 resistance in the Tamar Wallaby,
 RT Macropus eugenii.";
 RL Mol. Cell. Endocrinol. 119:169-174(1996).
 CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
 CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
 CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
 CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 CC NR3 SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: S83227; AAB49508.2; -
 DR HSSP: P06401; 1A2B.
 DR InterPro: IPR000536; Hormone_rec_11g.
 DR InterPro: IPR001628; zf-C4.

DR Pfam: PF00104; hormone_rec_1.
 DR Pfam: PF00105; zf-C4; 1.
 DR PROSITE: PS00031; NUCLEAR_RECEPTOR; PARTIAL.
 KW Receptor; transcription regulation; DNA-binding; Nuclear protein;
 KW Zinc-finger; Steroid-binding.
 FT NON_TER 1 1
 FT DNA_BIND <1 16 NUCLEAR RECEPTOR-TYPE.
 FT ZN_FING <1 11 C4-TYPE.
 FT DOMAIN 65 >180 STEROID-BINDING.
 FT NON_TER 180 180
 SQ SEQUENCE 180 AA; 20480 MW; D1618A89061C7BF3 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 180;
 Best Local Similarity 75.0%; Pred. No. 28;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGGRK 8
 I::IIII
 Db 15 GMVLGGRK 22

RESULT 7
 RCL1_ARATH STANDARD: PRT; 375 AA.
 ID AC 09C579;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable RNA 3'-terminal phosphate cyclase-like protein.
 DE A15622100.
 GN Arabidopsis thaliana (mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. COLUMBIA;
 RX MEDLINE=21016721; PubMed=11130714;
 RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
 RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
 RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
 RA Nakazaki N., Naito K., Okumura S., Shino S., Takeuchi C., Wada T.,
 RA Tanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
 RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston J.,
 RA Habermann K., Murray J., Johnson D., Kohling T., Nelson R.,
 RA Steneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
 RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
 RA Du H., Edwards J., Fryman J., Haekens B., Lamar E., Latreille P.,
 RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,
 RA Wagner-McPherson C., Woliam A., Yoakum M., Bell M., Dedhia N., J.,
 RA Pannell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
 RA Kirchhoff K., Toth K., King K., Bahret A., Miller B., Marra M.,
 RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
 RA Voiclaert G., Wambolt R., Duesterhoeft A., Stiekema W., Pohl T.,
 RA Entian K.-D., Terry N., Hartley N., Bent E., Johnson S.,
 RA Langham S.-A., McCullagh B., Robben J., Gymnopoulos B., Zimmermann W.,
 RA Ransperger U., Wecker H., Balke K., Weiler E., Peters S.,
 RA van Staveren M., Dirke W., Woolfman P., Klein Lankhorst R.,
 RA Welzenegger T., Bothe G., Rose M., Hauf J., Berner S., Hempel S.,
 RA Feldpausch M., Lamberth S., Villariol R., Gleien J., Ardiles W.,
 RA Berts O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,
 RA Schueller C., Zaccaria P., Mewes H.-W., Beyan M., Franz P.F.;
 RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 408:823-826(2000).
 CC -1- FUNCTION: DOES NOT HAVE CYCLASE ACTIVITY. PLAYS A ROLE IN 40S-
 CC RIBOSOMAL-SUBUNIT BIOGENESIS IN THE EARLY PRE-RNA PROCESSING
 CC STEPS AT SITES A0, A1 AND A2 THAT ARE REQUIRED FOR PROPER
 CC MATURATION OF THE 18S RNA (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE RNA 3'-TERMINAL CYCLASE FAMILY.

CC SUBFAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AL589883; CAC34503.1; -
 DR InterPro: IPR000228; RTC.
 DR Pfam: PF01137; RTC; 1.
 KW Nuclear protein.
 SQ SEQUENCE 375 AA; 40874 MW; 3C7D286D4386D5AF CRC64;
 OY 1 GTVLGK 7
 DB 75 GTVLGK 81
 RESULT 8
 PRGR_SHEEP STANDARD; PRT; 377 AA.
 AC Q28590;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Progesterone receptor (PR) (Fragment).
 GN PGR OR NR3C3.
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; ovis.
 NC NCB1_TaxID=9940;
 RN [1]
 RC STRAIN-ROMANOV; TISSUE-Uterus;
 RA Madiou T., Tiffoche C., le Gal F., Pelletier J., Theuclant M.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
 CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
 CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
 CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 CC NR3 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z66555; CAA91447.1; -
 DR HSSP: P06401; 1A28.
 DR InterPro: IPR000536; Hormone_rec_1lg.
 DR InterPro: IPR000128; Progesterone_receptor.
 DR InterPro: IPR001628; zf-C4.
 DR Pfam: PF00104; hormone_rec_1.
 DR Pfam: PF02161; Prog_receptor; 1.
 DR Pfam: PF00105; zf-C4; 1.
 DR SMART: SM00430; HOL1; 1.
 DR SMART: SM00399; znf-C4; 1.
 DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.

KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
 KW Zinc-finger; Steroid-binding.
 FT NON_TER 1
 FT DOMAIN 1
 FT DNA_BIND 18 83 MODULATING, PRO-RICH.
 FT ZN_FING 18 38 NUCLEAR RECEPTOR-TYPE.
 FT ZN_FING 54 78 C4-TYPE.
 FT DOMAIN 128 >377 C4-TYPE.
 FT NON_TER 377 377 STEROID-BINDING.
 SQ SEQUENCE 377 AA; 42904 MW; 3141B6587F7493C CRC64;
 OY 1 GTVLGK 8
 DB 82 GVLGK 89
 RESULT 9
 PRGR_CHICK STANDARD; PRT; 786 AA.
 AC P07812; Q90946;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Progesterone receptor (PR).
 GN PGR OR NR3C3.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NC NCB1_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88166640; PubMed=3443098;
 RA Gronemeyer H., Turcotte B., Quirin-Sticker C., Bocquel M.T.,
 RA Meyer M.E., Krozowski Z., Jeltsch J.M., Lerouge T., Garnier J.M.,
 RA Chambon P.;
 RT "The chicken progesterone receptor: sequence, expression and
 RT functional analysis." *EMBO J.* 6:3985-3994(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91042592; PubMed=3153474;
 RA Conneely O.M., Dobson A.D.W., Tsai M.-J., Beattie W.G., Toft D.O.,
 RA Hückaby C.S., Zarucki T., Schrader W.T., O'Malley B.W.;
 RT "Sequence and expression of a functional chicken progesterone
 RT receptor." *Mol. Endocrinol.* 1:517-525(1987).
 RN [3]
 RP SEQUENCE OF 128-164 FROM N.A.
 RX MEDLINE=86289413; PubMed=2426779;
 RA Conneely O.M., Sullivan W.P., Toft D.O., Birnhauser M., Cook R.G.,
 RA Maxwell B.L., Zarucki T., Greene G.L., Schrader W.T.,
 RA O'Malley B.W.;
 RT "Molecular cloning of the chicken progesterone receptor." *Science* 233:767-770(1986).
 RN [4]
 RP SEQUENCE OF 417-490 FROM N.A.
 RX MEDLINE=86287271; PubMed=2426697;
 RA Jeltsch J.M., Krozowski Z., Quirin-Sticker C., Gronemeyer H.,
 RA Simpson R.J., Garnier J.M., Krust A., Jacob F., Chambon P.;
 RT "Cloning of the chicken progesterone receptor." *Proc. Natl. Acad. Sci. U.S.A.* 83:5424-5428(1986).
 RN [5]
 RP DIFFERENCE BETWEEN FORM 1 AND FORM 2.
 RX MEDLINE=89340509; PubMed=2760059;
 RA Conneely O.M., Kettelberger D.M., Tsai M.-J., Schrader W.T.,
 RA O'Malley B.W.;
 RT "The chicken progesterone receptor A and B isoforms are products of

an alternate translation initiation event.";
 J. Biol. Chem. 264:14062-14064(1989).
 [6]
 MEDLINE-90154085; PubMed-2303468;
 J. Biol. Chem. 264:14062-14064(1989).
 J. Biol. Chem. 265:3967-3974(1990).
 FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
 THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
 PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
 SUBCELLULAR LOCATION: Nuclear.
 ALTERNATIVE PRODUCTS: 4 ISOFORMS: A (SHOWN HERE), A', B AND B';
 ARE PRODUCED BY ALTERNATIVE SPLICING.
 TISSUE SPECIFICITY: OVIDUCT.
 DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
 A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
 SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 NR3 SUBFAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

Query Match
 Best Local Similarity 78.0%; Score 32; DB 1; Length 786;
 Matches: 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 QY 1 CTIVGKK 8
 DB 485 GMVLGKK 492

 RESULT 10
 PRGR_MOUSE STANDARD; PRT; 923 AA.
 AC 000175;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Progesterone receptor (PR).
 GN PR OR NR3C3 OR PR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBI_Taxid=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=95100931; PubMed=7802637;
 RA Hagiwara K., Wu-Peng X.S., Funabashi T., Kato J., Pfaff D.W.;
 RT "Nucleic acid sequence and phase hypersensitive sites of the 5'
 RT region of the mouse progesterone receptor gene";
 RL Biochem. Biophys. Res. Commun. 205:1093-1101(1994).
 CC FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
 THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
 PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
 CC SUBCELLULAR LOCATION: Nuclear.
 CC DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
 CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
 CC SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 CC NR3 SUBFAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR HSP; P06401; 1A28.
DR MGI:97567; Paf.
DR InterPro: IPR000536; Hormone_rec_11g.
DR InterPro: IPR000128; Progesterone_receptor.
DR Pfam: PF00104; hormone_rec_1.
DR Pfam: PF02161; Progesterone_receptor.
DR Pfam: PF00105; zfc4; 1.
DR SMART: SM00430; HOL1; 1.
DR SMART: SM00399; znf_C4; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
DR Receptor: Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Steroid-binding.
FT DOMAIN 1 556 MODULATING, PRO-RICH.
FT DNA_BIND 557 622 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 557 577 C4-TYPE.
FT ZN_FING 593 617 C4-TYPE.
FT DOMAIN 671 923 STEROID-BINDING.
FT DOMAIN 184 188 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 923 AA; 99073 MW; 9415F1ED343BEEF CRC64;

Query Match 78.0%; Score 32; DB 1; Length 923;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGKK 8
Db 621 GMVLGKK 628
11111111

RESULT 11
ID PRGR_RAT STANDARD; PRT; 923 AA.
AC 063449;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Progesterone receptor (PR).
GN PGR OR NR3C3.
OS Rattus norvegicus (Rat).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CC NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Placenta;
RX MEDLINE=94130817; PubMed=8299566;
RA Park-Sarge O.K., Mayo K.E.,
RT "Regulation of the progesterone receptor gene by gonadotropins and
RT cyclic adenosine 3',5'-monophosphate in rat granulosa cells.";
RL Endocrinology 134:709-718(1994).
CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR3 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L16922; AAA19916.1; -
CC HSP; P06401; 1A28.
DR InterPro: IPR000536; Hormone_rec_11g.

DR InterPro: IPR000128; Progesterone_receptor.
DR InterPro: IPR001628; zfc4.
DR Pfam: PF00104; hormone_rec_1.
DR Pfam: PF02161; Progesterone_receptor.
DR Pfam: PF00105; zfc4; 1.
DR PRINTS: PR00047; STEROIDFINGER.
DR SMART: SM00430; HOL1; 1.
DR SMART: SM00399; znf_C4; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
DR Receptor: Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Steroid-binding.
FT DOMAIN 1 556 MODULATING, PRO-RICH.
FT DNA_BIND 557 622 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 557 577 C4-TYPE.
FT ZN_FING 593 617 C4-TYPE.
FT DOMAIN 671 923 STEROID-BINDING.
FT DOMAIN 184 188 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 923 AA; 99407 MW; 05384B9656BF22DC CRC64;

Query Match 78.0%; Score 32; DB 1; Length 923;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGKK 8
Db 621 GMVLGKK 628
11111111

RESULT 12
ID PRGR_RABIT STANDARD; PRT; 930 AA.
AC P06186;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Progesterone receptor (PR).
GN PGR OR NR3C3.
OS Oryctolagus cuniculus (Rabbit).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Euthera; Lagomorpha; Leporidae; Oryctolagus.
CC NCBI_Taxid=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87067449; PubMed=3538016;
RA Loofstiel H., Atger M., Mistrani M., Guiochon-Mantel A., Meriel C.,
RA Logeat F., Benarous R., Milgrom E.,
RT "Cloning and sequence analysis of rabbit progesterone-receptor
RT complementary DNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9045-9049(1986).
CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR3 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M14547; AAA14443.1; -
CC PIR; A25923; A25923.
DR HSP; P06401; 1A28.
DR TRANSFAC; T00697; -
DR InterPro: IPR000536; Hormone_rec_11g.
DR InterPro: IPR000128; Progesterone_receptor.

DR InterPro: IPR001628; zf-C4.
 DR Pfam: PF00104; hormone_rec. 1.
 DR Pfam: PF02161; Prog_receptor. 1.
 DR Pfam: PF00105; zf-C4. 1.
 DR PRINTS: PR00047; STEROIDINER.
 DR SMART: SM00430; HOLI. 1.
 DR SMART: SM00399; ZNF_C4. 1.
 DR PROSITE: PS00031; NUCLEAR_RECEPTOR. 1.
 DR Receptor: Transcription regulation; DNA-binding; Nuclear protein;
 zinc-finger; Steroid-binding.
 KW Receptor: Transcription regulation; DNA-binding; Nuclear protein;
 zinc-finger; Steroid-binding.
 FT DOMAIN 1 565 MODULATING, PRO-RICH.
 FT DNA_BIND 568 633 NUCLEAR RECEPTOR-TYPE.
 FT ZN_FING 568 588 C4-TYPE.
 FT ZN_FING 604 628 C4-TYPE.
 FT DOMAIN 678 930 STEROID-BINDING.
 FT DOMAIN 184 188 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 SO SEQUENCE 930 AA; 98666 MW; 644FFAC13BF2F883 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 930;
 Best Local Similarity 75.0%; Pred. No. 1.3e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GYVLGGRK 8
 DB 632 GMYLGGRR 639

RESULT 13
 PRGR_HUMAN STANDARD; PRT; 933 AA.
 AC P06401; Q90PF7;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Progesterone receptor (PR).
 GN PGR OR NR3C3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE-90228361; PubMed-2328727;
 RA Kastner P., Krust A., Turcotte B., Stropp U., Tora L., Gronemeyer H.,
 Chambon P.;
 RT "Two distinct estrogen-regulated promoters generate transcripts
 encoding the two functionally different human progesterone receptor
 forms A and B.";
 RT EMBO J. 9:1603-1614(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87184565; PubMed-3551956;
 RA Mistran M., Alger M., D'Aurion L., Loosfelt H., Meriel C.,
 Fridlansky F., Guichon-Mantel A., Gilbert F., Milgrom E.;
 RT "Complete amino acid sequence of the human progesterone receptor
 deduced from cloned cDNA.";
 RT Biochem. Biophys. Res. Commun. 143:740-748(1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Kieback D.G., Agoulrik I.U., Tong X.-W.;
 RL Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 682-933.
 RA Williams S.P., Sigler P.B.;
 RT "Atomic structure of progesterone complexed with its receptor.";
 RL Nature 393:392-396(1998).
 CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
 CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
 CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE

CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
 CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 CC NR3 SUBFAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use, by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X51730; CAA36018.1; -;
 DR EMBL: M15716; AAA60081.1; -;
 DR EMBL: AF016381; AAD01587.1; -;
 DR PIR: A03245; QRHP.
 DR PIR: S09971; S09971.
 DR PDB: 1A28; 15-JUL-98.
 DR TRNSPAC: T00696; -;
 DR MIM: 264080; -;
 DR InterPro: IPR000536; Hormone_rec_1lg.
 DR InterPro: IPR000128; Progesterone_receptor.
 DR InterPro: IPR001723; Steroidhormone_receptor.
 DR InterPro: IPR001628; zf-C4.
 DR Pfam: PF00104; hormone_rec. 1.
 DR Pfam: PF02161; Prog_receptor. 1.
 DR PRINTS: PR000396; STROHORMONER.
 DR PRINTS: PR00047; STROIDFINGER.
 DR SMART: SM00430; HOLI. 1.
 DR SMART: SM00399; ZNF_C4. 1.
 DR PROSITE: PS00031; NUCLEAR_RECEPTOR. 1.
 KW Receptor: Transcription regulation; DNA-binding; Nuclear protein;
 zinc-finger; Steroid-binding; Phosphorylation; Alternative splicing;
 3D-structure.
 FT DOMAIN 1 566 MODULATING, PRO-RICH.
 FT DNA_BIND 567 632 NUCLEAR RECEPTOR-TYPE.
 FT ZN_FING 567 587 C4-TYPE.
 FT ZN_FING 603 627 C4-TYPE.
 FT DOMAIN 681 933 STEROID-BINDING.
 FT DOMAIN 183 187 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT MOD_RES 41 41 PHOSPHORYLATION (BY CK).
 FT MOD_RES 227 227 PHOSPHORYLATION (BY CK).
 FT MOD_RES 232 232 PHOSPHORYLATION (BY CK).
 FT MOD_RES 552 552 PHOSPHORYLATION (BY CK).
 FT MOD_RES 793 793 PHOSPHORYLATION (BY CK).
 FT VARSPIC 1 164 MISSING (IN ISOFORM B).
 FT CONFLICT 226 226 G -> S (IN REF. 1).
 FT CONFLICT 256 256 V -> S (IN REF. 1).
 FT CONFLICT 344 344 S -> T (IN REF. 1).
 FT CONFLICT 660 660 L -> V (IN REF. 2).
 SO SEQUENCE 933 AA; 98994 MW; 80414B7F1317F8E CRC64;

Query Match 78.0%; Score 32; DB 1; Length 933;
 Best Local Similarity 75.0%; Pred. No. 1.3e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GYVLGGRK 8
 DB 631 GMYLGGRR 638

RESULT 14
 ID YOKX_PYRMO STANDARD; PRT; 148 AA.
 AC P20298;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein in GAPDH 3' region (ORF X) (Fragment).

OS Pyrococcus woesei.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 RN NCBI_TaxID=2262;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSM 3773;
 RX MEDLINE=90330536; PubMed=2165475;
 RA Zwickl P., Fabry S., Bogedain C., Haas A., Hensel R.;
 RT "Glyceroldehyde-3-phosphate dehydrogenase from the hyperthermophilic
 archaeobacterium Pyrococcus woesei: characterization of the enzyme,
 RT cloning and sequencing of the gene, and expression in Escherichia
 coli.";
 RL J. Bacteriol. 172:4329-4338(1990).
 DR PIR: S10655; S10655.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 148 AA; 15256 MW; 8C7E51A072DF5D50 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 148;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GTVLCG 6
 |||||
 Db 89 GTVLCG 94

RESULT 15
 CATB_RAT STANDARD; PRT; 339 AA.
 AC P00787;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Cathepsin B precursor (EC 3.4.22.1) (Cathepsin B1) (RSG-2).
 CN CTGB.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-Mammary gland;
 RX MEDLINE=95094788; PubMed=8001549;
 RA Guenette R.S., Woolbroek M., Wong K., Wong P., Tenniwood M.;
 RT "Cathepsin B, a cysteine protease implicated in metastatic
 RT progression, is also expressed during regression of the rat prostate
 RT and mammary glands.";
 RL Eur. J. Biochem. 226:311-321(1994).
 RN [2]
 RP SEQUENCE OF 69-339 FROM N.A.
 RX MEDLINE=85190489; PubMed=2986112;
 RA San Segundo B., Chan S.J., Steiner D.F.;
 RT "Identification of cDNA clones encoding a precursor of rat liver
 RT cathepsin B.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2320-2324(1985).
 RN [3]
 RP SEQUENCE OF 80-126 AND 129-333.
 RC TISSUE-Liver;
 RX MEDLINE=83221657; PubMed=6574504;
 RA Takio K., Towatari T., Katunuma N., Teller D.C., Titani K.;
 RT "Homology of amino acid sequences of rat liver cathepsins B and H
 RT with that of papain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:3666-3670(1983).
 RN [4]
 RP PROCESSING.
 RX MEDLINE=92348471; PubMed=1639824;
 RA Rowan A.D., Mason P., Mach L., Mort J.S.;
 RT "Rat procathepsin B. Proteolytic processing to the mature form in
 RT vitro.";
 RL J. Biol. Chem. 267:15993-15999(1992).
 RN [5]

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE=95197558; PubMed=7890671;
 RA Uta Z., Hasnain S., Hirama T., Lee X., Mort J.S., To R., Huber C.P.;
 RT "Crystal structures of recombinant rat cathepsin B and a cathepsin B-
 RT inhibitor complex. Implications for structure-based inhibitor
 RT design.";
 RL J. Biol. Chem. 270:5527-5533(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 18-339.
 RX MEDLINE=96311276; PubMed=8740363;
 RA Cygler M., Sivaraman J., Grochulski P., Coulombe R., Storer A.C.,
 RA Mort J.S.;
 RT "Structure of rat procathepsin B: model for inhibition of cysteine
 RT protease activity by the proregion.";
 RL Structure 4:405-416(1996).
 CC -1- FUNCTION: THIOL PROTEASE WHICH IS BELIEVED TO PARTICIPATE IN
 CC INTRACELLULAR DEGRADATION AND TURNOVER OF PROTEINS. HAS ALSO
 CC BEEN IMPLICATED IN TUMOR INVASION AND METASTASIS.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSES PROTEINS, WITH A SPECIFICITY
 CC RESEMBLING THAT OF PAPAIN.
 CC -1- SUBUNIT: DIMER OF A HEAVY CHAIN AND A LIGHT CHAIN CROSS-LINKED
 CC BY A DISULFIDE BOND.
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
 CC PAPAIN FAMILY OF THIOL PROTEASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).

CC EMBL: X82396; CAA57792.1; -
 CC EMBL: M1305; AAA40993.1; -
 CC PIR: A00977; KHRTB.
 DR PDB: 1CPT; 07-DEC-95.
 DR PDB: 1MR; 11-JAN-97.
 DR PDB: 1CTE; 31-JUL-95.
 DR PDB: 1THE; 10-MAR-96.
 DR MEROPS: C01.060; -
 DR Interpro: IPR000668; Peptidase_C1.
 DR Interpro: IPR000169; Thiolprot_act_site.
 DR Pfam: PF00112; Peptidase_C1; 1.
 DR PRINTS: PR00705; PAPAIN.
 DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS; 1.
 DR PROSITE: PS00640; THIOL_PROTEASE_ASN; 1.
 DR Hydrolyse: Thiol protease; Lysosome; Glycoprotein; Zymogen; Signal;
 KW 3D-structure.
 KM SIGNAL 1 17
 FT PROPEP 18 79 POTENTIAL.
 FT CHAIN 80 333 ACTIVATION PEPTIDE.
 FT CHAIN 80 126 CATHEPSIN B.
 FT CHAIN 129 333 CATHEPSIN B LIGHT CHAIN.
 FT PROPEP 334 339 CATHEPSIN B HEAVY CHAIN.
 FT ACT_SITE 108 108
 FT ACT_SITE 278 278
 FT ACT_SITE 298 298
 FT DISULFID 93 122 BY SIMILARITY.
 FT DISULFID 105 150 BY SIMILARITY.
 FT DISULFID 141 207 BY SIMILARITY.
 FT DISULFID 142 146 BY SIMILARITY.
 FT DISULFID 179 211 BY SIMILARITY.
 FT DISULFID 187 198 BY SIMILARITY.
 FT CARBOHYD 192 192
 FT VARIANT 302 302
 FT CONFLICT 159 159
 FT CONFLICT 159 159
 SQ SEQUENCE 339 AA; 37470 MW; 925E2E36C2B03CDA CRC64;

Query Match 75.6%; Score 31; DB 1; Length 339;

Best Local Similarity 100.0%; Pred. NO. 79;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTVGG 6
|||||
Db 60 GTVGG 65

Search completed: July 30, 2002, 15:37:59
Job time: 313 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:37:26 ; Search time 88.56 Seconds

(Without alignments)
15.627 Million cell updates/sec

Title: US-09-813-214A-2

Perfect score: 41

Sequence: 1 GTVLGSKK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_ARCHAE:*

2: SP_BACTERIA:*

3: SP_FUNGI:*

4: SP_HUMAN:*

5: SP_INVERTEBRATE:*

6: SP_MAMMAL:*

7: SP_MNC:*

8: SP_ORGANELLE:*

9: SP_PLANT:*

10: SP_PROTOZOA:*

11: SP_TODONT:*

12: SP_VIRUS:*

13: SP_VERTEBRATE:*

14: SP_UNCLASSIFIED:*

15: SP_VIRUS:*

16: SP_BACTERIAP:*

17: SP_ARCHAE:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 37 | 90.2 | 525 | 3 | 09C214 |
| 2 | 35 | 85.4 | 251 | 16 | 09JWL1 |
| 3 | 35 | 85.4 | 286 | 11 | 09DCV0 |
| 4 | 35 | 85.4 | 509 | 2 | 09X6P3 |
| 5 | 34 | 82.9 | 214 | 16 | 09PEW0 |
| 6 | 33 | 80.5 | 193 | 10 | 09LJG6 |
| 7 | 33 | 80.5 | 227 | 16 | 09LJG6 |
| 8 | 33 | 80.5 | 244 | 16 | 09LJG6 |
| 9 | 33 | 80.5 | 537 | 5 | 09Y154 |
| 10 | 33 | 80.5 | 545 | 5 | 09Y154 |
| 11 | 33 | 80.5 | 607 | 13 | 09PS81 |
| 12 | 33 | 80.5 | 689 | 10 | 09C9J3 |
| 13 | 32 | 78.0 | 64 | 10 | 09FP75 |
| 14 | 32 | 78.0 | 153 | 11 | 09Z3G6 |
| 15 | 32 | 78.0 | 174 | 12 | 09WHH2 |
| 16 | 32 | 78.0 | 174 | 12 | 09PZ70 |

| | | | | | |
|----|----|------|-------|----|--------|
| 17 | 32 | 78.0 | 179 | 16 | 09A1L4 |
| 18 | 32 | 78.0 | 254 | 2 | 09Z445 |
| 19 | 32 | 78.0 | 288 | 16 | 09Z546 |
| 20 | 32 | 78.0 | 288 | 6 | 028547 |
| 21 | 32 | 78.0 | 337 | 16 | 097LX2 |
| 22 | 32 | 78.0 | 349 | 16 | 09ZNV6 |
| 23 | 32 | 78.0 | 360 | 13 | 042274 |
| 24 | 32 | 78.0 | 375 | 16 | 09RYA6 |
| 25 | 32 | 78.0 | 376 | 4 | 09H119 |
| 26 | 32 | 78.0 | 454 | 10 | 09FX33 |
| 27 | 32 | 78.0 | 496 | 16 | 09HW92 |
| 28 | 32 | 78.0 | 583 | 13 | 09DEV4 |
| 29 | 32 | 78.0 | 717 | 10 | 09T045 |
| 30 | 32 | 78.0 | 732 | 13 | 09DDU9 |
| 31 | 32 | 78.0 | 939 | 6 | 09G1M0 |
| 32 | 32 | 78.0 | 960 | 5 | 09VTK6 |
| 33 | 32 | 78.0 | 1078 | 10 | 094FR5 |
| 34 | 32 | 78.0 | 1127 | 12 | 036360 |
| 35 | 32 | 78.0 | 10917 | 2 | 093NM6 |
| 36 | 32 | 78.0 | 11096 | 2 | 091AW3 |
| 37 | 31 | 75.6 | 77 | 16 | 098RT3 |
| 38 | 31 | 75.6 | 99 | 15 | 098117 |
| 39 | 31 | 75.6 | 108 | 3 | 013531 |
| 40 | 31 | 75.6 | 110 | 11 | 09JK77 |
| 41 | 31 | 75.6 | 140 | 11 | 09EPC6 |
| 42 | 31 | 75.6 | 149 | 11 | 09JHU7 |
| 43 | 31 | 75.6 | 157 | 16 | 09XEQ1 |
| 44 | 31 | 75.6 | 159 | 16 | 097D00 |
| 45 | 31 | 75.6 | 175 | 16 | 055671 |

ALIGNMENTS

RESULT 1

09C214

ID 09C214 PRELIMINARY: PRT: 525 AA.

AC 09C214:

DT 01-JUN-2001 (Trembl, 17, Created)

DT 01-DEC-2001 (Trembl, 19, Last sequence update)

DT 01-DEC-2001 (Trembl, 19, Last annotation update)

DE RELATED TO DNA POLYMERASE DELTA SMALL SUBUNIT.

GN 93611.080.

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariales; Sordariaceae; Neurospora.

OX NCBI_taxid=5141;

RN [1]

RP SEQUENCE FROM N.A.

RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,

RA Nykattura G., Mewes H.W., Mannhaupt G.;

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA German Neurospora genome project;

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL, AL513443; CAC28660.2; -

SQ SEQUENCE 525 AA; 58146 MW; 7FF76E8B05C66F04 CRC64;

Query Match 90.2%; Score 37; DB 3; Length 525;
Best local similarity 87.5%; Pred. No. 48;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

09 1 GTVLGSKK 8

DB 74 GTVLGSKK 81

RESULT 2

09JWL1 PRELIMINARY: PRT: 251 AA.

AC 09JWL1:

DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE PUTATIVE MAFB ALTERNATIVE C-TERMINUS (FRAGMENT).
 GN MAFB OR NMA0319.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=56699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,
 RA Jagsels K., Leacher S., Moule S., Mungall K., Quail M.A.,
 RA Rajadaram M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrall B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis Z2491."
 RL Nature 404:502-506(2000).
 DR EMBL: AL162752; CAB83624.1; -.
 KW Complete proteome.
 FT NON-TER
 SQ SEQUENCE 251 AA; 27783 MW; E6E47DF3CEDF2C7C CRC64;

Query Match 85.4%; Score 35; DB 16; Length 251;
 Best Local Similarity 87.5%; Pred. No. 54;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGGRK 8
 Db 68 GTVMGGRK 75

RESULT 3
 ID Q9DCV0 PRELIMINARY; PRT; 286 AA.
 AC Q9DCV0;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE ADULT MALE KIDNEY CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
 DE CLONE:0610010D10, FULL INSERT SEQUENCE.
 GN CTSS.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
 RA Kuenl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubl F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Mysnha-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."

RL Nature 409:685-690(2001).
 DR EMBL: AK002456; BAB22113.1; -.
 DR HSSP: P00787; ITHE.
 DR MGD: MGI:88561; CtSD.
 DR InterPro: IPR000668; Peptidase_C1.
 DR InterPro: IPR000169; Thiolprot_act_site.
 DR Pfam: PF00112; Thiolactase_C1; 1.
 DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
 KW Hydrolyase, thiol protease.
 SQ SEQUENCE 286 AA; 31091 MW; CB9F045097126011 CRC64;

Query Match 85.4%; Score 35; DB 11; Length 286;
 Best Local Similarity 87.5%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGGRK 8
 Db 60 GTVLGGRK 67

RESULT 4
 ID Q9X6P3 PRELIMINARY; PRT; 509 AA.
 AC Q9X6P3;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE ADHESIN MAFB.
 GN MAFB.
 OS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=485;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MS11;
 RA Eickernjaeger S., Meyer T.F., Fischer E., Maier J., Manning P.A.,
 RA Rudel T., Schuenepflug I., Schulz E., Schwan E.T.;
 RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF142582; AAD31039.1; -.
 SQ SEQUENCE 509 AA; 55681 MW; 88B420CFED8269D4 CRC64;

Query Match 85.4%; Score 35; DB 2; Length 509;
 Best Local Similarity 87.5%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGGRK 8
 Db 326 GTVMGGRK 333

RESULT 5
 ID Q9PEW0 PRELIMINARY; PRT; 214 AA.
 AC Q9PEW0;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE ACYL-[ACP]-UDP-N-ACETYLGLUCOSAMINE.
 GN XPG918.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,

RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Fetto J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frome M., Furian L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Holsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Silvestri W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchiko M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.,
 RA "The genome sequence of the plant pathogen *Xylella fastidiosa*,"
 RT Nature 406:151-159(2000).
 RL EMBL: AE003931; AAF83728.1;
 DR InterPro: IPR001451; Hexapep_transf.
 DR Pfam: PF00132; hexapep; 9.
 DR PROSITE: PS00101; HEXAPEP_TRANSFERASES; UNKNOWN-3.
 KW Complete proteome.
 SQ SEQUENCE 214 AA; 22275 MW; 57AE135939845397 CRC64;

Query Match 82.9%; Score 34; DB 16; Length 214;
 Best Local Similarity 85.7%; Pred. No. 71;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GTVIGGK 7
 Db 59 GTVIGGK 65

RESULT 6
 ID 09LJG6 PRELIMINARY; PRT: 193 AA;
 AC 09LJG6;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
 DE GENOMIC DNA, CHROMOSOME 3, P1 CLONE: MAG2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.,
 RA Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-20363099; PubMed=10907853;
 RX MEDLINE-20363099; PubMed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 DR EMBL: AP000600; BAB02986.1;
 SQ SEQUENCE 193 AA; 21732 MW; DC8718A166F75161 CRC64;

Query Match 80.5%; Score 33; DB 10; Length 193;

Best Local Similarity 87.5%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GTVIGGK 8
 Db 34 GTVIGGK 41
 RESULT 7
 ID 092G42 PRELIMINARY; PRT: 227 AA;
 AC 092G42;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
 DE HYPOTHETICAL PROTEIN RC1282.
 GN RC1282.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OC NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MALISH 7;
 RX MEDLINE-21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii,"
 RL Science 293:2093-2098(2001).
 DR EMBL: AE008675; AAL03820.1;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 227 AA; 24449 MW; 315196E84D5E2992 CRC64;

Query Match 80.5%; Score 33; DB 16; Length 227;
 Best Local Similarity 71.4%; Pred. No. 1.2e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GTVIGGK 7
 Db 82 GTVIGGK 88
 RESULT 8
 ID 092XC7 PRELIMINARY; PRT: 244 AA;
 AC 092XC7;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
 DE HYPOTHETICAL PROTEIN SMO20032.
 GN SMO20032.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Plasmid symb (megaplasmid 2).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OC NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-1021;
 RX MEDLINE-21396508; PubMed=11481431;
 RA Finn T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
 RA Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Guzy J.,
 RA Golding B., Pehler A.;
 RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
 RT fixing endosymbiont Sinorhizobium meliloti,"
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
 DR EMBL: AL603642; CAC48435.1;
 KW Plasmid; Hypothetical protein; Complete proteome.
 SQ SEQUENCE 244 AA; 26022 MW; 848748E827BA876A CRC64;

Query Match

80.5%; Score 33; DB 16; Length 244;

Best Local Similarity 85.7%; Pred. No. 1.3e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTVIGSK 7

Db 214 GTLIGSK 220

RESULT 9
ID 09Y154 PRELIMINARY; PRT; 537 AA.
AC 09Y154;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE BCDA:GH03163 OR CG1408.
GN BCDA:GH03163 OR CG1408.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Anil J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burdits K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).

RA Celniker S.E.;
RT "Full-length Drosophila melanogaster cDNA sequence."
RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003775; AAF57089.1;
DR EMBL; AF145614; AAD38589.1;
DR Flybase; FBgn0027598; BCDA:GH03163.
SQ SEQUENCE 537 AA; 56128 MW; 704755F0092246B6 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 537;
Best Local Similarity 75.0%; Pred. No. 3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVIGSK 8
Db 175 GNLLGSK 182

RESULT 10
ID 09VA35 PRELIMINARY; PRT; 545 AA.
AC 09VA35;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE BCDA:GH03163 OR CG1408.
GN BCDA:GH03163 OR CG1408.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Anil J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burdits K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).
 DR EMBL: AF003775; AAF57088.1; -
 DR FlyBase: FBgn0027598; BCDNA:GH03163
 SQ SEQUENCE 545 AA; 57302 MW; DBB92C9B979BF768 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 545;
 Best Local Similarity 75.0%; Pred. NO. 3.1e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGSKK 8
 ||:||||
 Db 175 GNVLGSKK 182

RESULT 11
 ID 09PS81 PRELIMINARY; PRT; 607 AA.
 AC 09PS81:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE MAD+:PROTEIN(ADP-RIBOSYL)-TRANSFERASE, ADPRT
 OS Xenopus.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 OC Xenopodinae.
 OC NCBI_TaxID=8353;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96007847; PubMed=7578427;
 RA Auer B., Flick K., Wang Z.Q., Haidacher D., Jager S., Berghammer H.,
 RA Kofler B., Schweiger M., Wagner E.F.;
 RA "On the biological role of the nuclear polymerizing NAD+: protein(ADP-
 RT ribosyl) transferase (ADPRT): ADPRT from Dictyostelium discoideum and
 RT inactivation of the ADPRT gene in the mouse.";
 RL Biochimie 77:444-449(1995).
 DR HSSP: P26446; 1A26.
 DR InterPro: IPR001357; BRCT.
 DR InterPro: IPR001290; PARP.
 DR InterPro: IPR004102; PARP_reg.
 DR Pfam: PF00533; BRCT; 1.
 DR Pfam: PF00644; PARP; 1.
 DR Pfam: PF02877; PARP_reg; 1.
 DR PROSITE: PS0172; BRCT; 1.
 SQ SEQUENCE 607 AA; 67496 MW; 54CDEBBE22079886 CRC64;

Query Match 80.5%; Score 33; DB 13; Length 607;
 Best Local Similarity 75.0%; Pred. NO. 3.5e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGSKK 8
 ||:||||
 Db 185 GTVLGSKK 192

RESULT 12
 ID 09C913 PRELIMINARY; PRT; 689 AA.
 AC 09C913:
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL 77.5 KDA PROTEIN.
 GN F26A9.16.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Eumariophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
 RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,
 RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome 1 BAC F26A9 sequence.";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC016163; AAG51819.1;
 DR InterPro: IPR002885; PPR.
 DR Pfam: PF01535; PPR; 6.
 KW Hypothetical protein.
 SQ SEQUENCE 689 AA; 77486 MW; 16413082FF846D0D CRC64;

Query Match 80.5%; Score 33; DB 10; Length 689;
 Best Local Similarity 75.0%; Pred. NO. 4e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGSKK 8
 ||:||||
 Db 186 GTVLGSKK 193

RESULT 13
 ID 09EP75 PRELIMINARY; PRT; 64 AA.
 AC 09EP75:
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE P0458A05.14 PROTEIN (B1157F09.4 PROTEIN).
 GN P0458A05.14 OR B1157F09.4.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Elmhartoideae; Oryzaceae; Oryza.
 OC NCBI_TaxID=4530;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0458A05.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
 RT clone:B1157F09.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP002870; BAB19405.1;
 DR EMBL: AP003207; BAB64079.1;
 SQ SEQUENCE 64 AA; 7300 MW; 269844DAF129933B CRC64;

Query Match 78.0%; Score 32; DB 10; Length 64;
 Best Local Similarity 75.0%; Pred. NO. 48;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTVLGSKK 8
 ||:||||
 Db 53 GTVLGSKK 60

RESULT 14
 ID 0923G6 PRELIMINARY; PRT; 153 AA.
 AC 0923G6:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PROGESTERONE RECEPTOR (FRAGMENT).
 GN PR.

Search completed: July 30, 2002, 15:37:28
Job time: 333 sec

OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang X., Das S.K., Paria B.C.;
RT "Mesocricetus auratus progesterone receptor (PR) mRNA."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY040534; AAK7950.1; -.
KW Receptor.
FT NON_TER 1 1
FT NON_TER 153 153
SQ SEQUENCE 153 AA; 16909 MW; C0F00BF4ED3B8D95 CRC64;

Query Match 78.0%; Score 32; DB 11; Length 153;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGKK 8
| | | | | | | |
DB 87 GTVLGKK 94

RESULT 15

O9WHH2 PRELIMINARY; PRT; 174 AA.
AC O9WHH2;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE M MATRIX PROTEIN.
OS Porcine reproductive and respiratory syndrome virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID=28344;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH-1A;
RA Tong G.-Z., Qiu H.-J., Zhou Y.-J., Guo B.-Q., Zhang S.-J., Wang L.,
CAI X.-H.;
RT "Molecular Analysis of Genes encoding for the Structural Proteins of
RT Porcine Reproductive and Respiratory Syndrome Virus Strain CH-1A."
RL Ziran kexue jinzhan 10:147-153(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CH-1A;
RA Tong G.-Z., Qiu H.-J., Liu G.-Q., Xue Q., Zhou Y.-J., Qian P.,
Guo B.-Q., Liu B.-Q.;
RT "Complete Genomic Sequence of Porcine Reproductive and Respiratory
RT Syndrome Virus Strain CH-1A."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY032626; AAK44217.1; -.
DR InterPro: IPR001332; Arter1_glycop.
DR InterPro: IPR003240; Porcine_RR_virus_M.
DR Pfam: PF00951; Arter1_glycop; 1.
DR ProDom: PD002878; Porcine_RR_virus_M; 1.
KW Matrix protein.
SQ SEQUENCE 174 AA; 19011 MW; 69B81EBB511712BB CRC64;

Query Match 78.0%; Score 32; DB 12; Length 174;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTVLGKK 8
| | | | | | | |
DB 153 GTVLGKK 160

Wed Jul 31 07:54:09 2002

us-09-813-214a-2.rsp

1
2
3

1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:34:12 ; Search time 82.85 Seconds

(without alignments)
2844.883 Million cell updates/sec

Title: US-09-813-214A-9

Perfect score: 10708
Sequence: 1 MNHIVYVIFNKATGTFMAYV.....NCSADTQGHVCAVAGGFHF 2122

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A.GeneSeq_032802:*
1: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneSeq/geneSeq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|---------|-------------|--------|-------|-------------|
| 1 | 10697.5 | 99.9 | 2123 | 22 | AAE00701 |
| 2 | 7380.5 | 68.9 | 2047 | 22 | AAB69134 |
| 3 | 7357.5 | 68.7 | 2053 | 22 | AAB69135 |
| 4 | 7113.5 | 66.4 | 1992 | 22 | AAB69133 |
| 5 | 7110.5 | 66.4 | 1992 | 17 | AAW04505 |
| 6 | 7110.5 | 66.4 | 1992 | 22 | AAB69137 |
| 7 | 6600 | 61.6 | 2314 | 22 | AAB69136 |
| 8 | 1185.5 | 11.1 | 2353 | 17 | AA993933 |
| 9 | 1169.5 | 10.9 | 2411 | 22 | AA823860 |
| 10 | 1063 | 9.9 | 1778 | 22 | ABB52677 |
| 11 | 1018.5 | 9.5 | 1104 | 21 | AA823856 |

| | | | | | | |
|----|--------|-----|------|----|----------|---------------------|
| 12 | 1018.5 | 9.5 | 1104 | 21 | AA823859 | Haemophilus influe |
| 13 | 974 | 9.1 | 1004 | 21 | AA823857 | Haemophilus influe |
| 14 | 944.5 | 8.8 | 1002 | 21 | AA823854 | Haemophilus influe |
| 15 | 809.5 | 7.6 | 2039 | 19 | AA856322 | Haemophilus paraga |
| 16 | 723 | 6.8 | 2042 | 19 | AA856319 | Haemophilus paraga |
| 17 | 693 | 6.5 | 1094 | 21 | AA823858 | Haemophilus influe |
| 18 | 681 | 6.4 | 1098 | 17 | AA893932 | Haemophilus adhesi |
| 19 | 609.5 | 5.7 | 2344 | 22 | AA837120 | Staphylococcus aur |
| 20 | 601 | 5.6 | 5795 | 22 | AA837017 | Staphylococcus aur |
| 21 | 585.5 | 5.6 | 6281 | 22 | AA837403 | Staphylococcus aur |
| 22 | 582.5 | 5.4 | 679 | 17 | AA893934 | Haemophilus adhesi |
| 23 | 582.5 | 5.4 | 679 | 21 | AA823855 | Haemophilus influe |
| 24 | 550 | 5.1 | 2086 | 22 | AA834143 | Staphylococcus aur |
| 25 | 549.5 | 5.1 | 2586 | 22 | ABB66878 | Drosophila melanog |
| 26 | 543 | 5.1 | 2434 | 22 | AA834339 | Staphylococcus aur |
| 27 | 539.5 | 5.0 | 2539 | 21 | AA875098 | Staphylococcus aur |
| 28 | 537 | 5.0 | 1601 | 18 | AA830292 | Neisseria meningit |
| 29 | 533.5 | 5.0 | 2732 | 22 | AA852855 | Non-tyeable Haemo |
| 30 | 528.5 | 4.9 | 1598 | 18 | AA830291 | Escherichia coli p |
| 31 | 522.5 | 4.9 | 2478 | 22 | AA834320 | Non-tyeable Haemo |
| 32 | 522.5 | 4.9 | 2478 | 22 | AA837374 | Staphylococcus aur |
| 33 | 515 | 4.8 | 1536 | 14 | AA841725 | High molecular wei |
| 34 | 513 | 4.8 | 2514 | 21 | AA875097 | Neisseria meningit |
| 35 | 512 | 4.8 | 1536 | 15 | AA863505 | Haemophilus high m |
| 36 | 512 | 4.8 | 1536 | 21 | AA801846 | Haemophilus influe |
| 37 | 510.5 | 4.8 | 2368 | 22 | AA834139 | Staphylococcus aur |
| 38 | 510.5 | 4.8 | 2368 | 22 | AA836796 | Staphylococcus aur |
| 39 | 509 | 4.8 | 1536 | 14 | AA841723 | High molecular wei |
| 40 | 508 | 4.7 | 1536 | 18 | AA830293 | Non-tyeable Haemo |
| 41 | 508 | 4.7 | 2383 | 21 | AA815945 | E. coli proliferat |
| 42 | 497 | 4.6 | 1529 | 14 | AA841732 | High molecular wei |
| 43 | 496.5 | 4.6 | 2048 | 21 | AA875096 | Neisseria gonorrhoe |
| 44 | 490.5 | 4.6 | 1477 | 14 | AA841724 | High molecular wei |
| 45 | 478.5 | 4.5 | 1477 | 15 | AA863506 | Haemophilus high m |

ALIGNMENTS

| | |
|----------|--|
| RESULT 1 | |
| AAE00701 | AAE00701 standard; Protein: 2123 AA. |
| ID | AAE00701 |
| XX | AAE00701; |
| AC | AAE00701; |
| XX | |
| DT | 02-JUL-2001 (first entry) |
| XX | |
| DE | Moraxella catarrhalis outer membrane protein-106 (OMP106). |
| XX | |
| KW | Outer membrane protein-106; OMP106; haemagglutination; vaccine; |
| KW | bacterial infection; immunogen; cytotoxic; antibiotic; |
| KW | passive immunisation; |
| XX | |
| OS | Moraxella catarrhalis. |
| XX | |
| PN | US6214981-B1. |
| XX | |
| PD | 10-APR-2001. |
| XX | |
| PF | 12-NOV-1997; 97US-0968685. |
| XX | |
| PR | 03-MAY-1996; 96US-0642712. |
| XX | |
| PA | (ANTE-) ANTEX BIOLOGICS INC. |
| XX | |
| PI | Tucker K, Plossia L, Tillman UF; |
| XX | |
| DR | WPI: 2001-281002/29; |
| XX | |
| DR | N-PSDB: AAD04029. |
| XX | |
| PT | Novel nucleotide sequences encoding Moraxella catarrhalis outer |
| PT | membrane protein-106 polypeptide; useful for diagnosis of bacterial |
| PT | infections and as vaccine against Moraxella catarrhalis infection of |

CC 1800 PAEAIIDRINEGIRFEHFHNDNOEPVVOGRGICDSSASGKHSVAIGFOAKDGEAAYAIG 1859
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 1801 paeaidrinegiri f h v n d n o e p v v o g r g i c d s s a s g k h s v a i g f o a k d g e a a y a i g 1860
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 1860 ROTOGANSIAGDMAATGQOSIAIGTGNVAGKHSAGIADPSTVKADNSGYVGNNOF 1919
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 1861 rltgagngsiagdndagtdgslatignvvaqkhsagidpstvkaadnsyvgnnmf 1920
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 1920 TDAQTDFEGVGNNTTYESNSVALGSNSAISAGTHAGTOAKKSDGTAGTTTATGATGV 1979
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 1921 tdaqtdfvgygnntt yes n s v a l g s n s a i s a g t h a g t o a k k s d g t a g t t t a t g a t g v 1980
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 1980 KGFAGQTVAGVAGSASAEKRIQVNAAGEVSATSTDAVNSOLYKATQIGANNTELDH 2039
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 1981 kfgagqtvagvagsasaeerriqnvagevsatstdavpysqlykatqiganna tne l d h 2040
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 2040 RIHONENKANGISSAMAMAPQAYIPGRSMVTGATFHNGOGAVAVGSLSDNGOMV 2099
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 2041 rihonenkanagissamama smpqay i p g r s m v t g a t f h n g o g a v a v g s l s d n g o m v 2100
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 2100 FKINGSADTQGHVGAAGAGFHF 2122
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 2101 fking sad t q g h v g a a g a g f h f 2123
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC RESULT 2
 CC AAB69134 standard; Protein: 2047 AA.
 CC ID AAB69134
 CC XX AAB69134;
 CC AC AAB69134;
 CC XX 24-APR-2001 (first entry)
 CC DT 24-APR-2001 (first entry)
 CC DE M. catarrhalis strain 4223 genomic 200kda protein seq ID NO:7.
 CC XX
 CC KM Moraxella catarrhalis strain 4223; major outer membrane protein;
 CC KM 200kda outer membrane protein; antibacterial; immunogenic; infection;
 CC KM otitis media; detection.
 CC XX
 CC OS Moraxella catarrhalis.
 CC XX
 CC PM WO200107619-A1.
 CC XX
 CC PD 01-FEB-2001.
 CC XX
 CC XX 26-JUL-2000; 2000WO-CA00870.
 CC PF 26-JUL-2000; 2000WO-CA00870.
 CC XX
 CC PR 27-JUL-1999; 99US-0361619.
 CC XX
 CC PA (CONN-) CONNAUGHT LAB LTD.
 CC XX
 CC PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
 CC XX
 CC DR WPI: 2001-159722/16.
 CC DR N-PEDB; AAF59102; AAF59103.
 CC XX
 CC XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
 CC PT useful in protective vaccines and for diagnosis
 CC XX
 CC PS Claim 1: Fig 3A-W; 247pp; English.
 CC XX
 CC XX The present invention describes an isolated and purified nucleic acid (I)
 CC CC that encodes a 200 kda outer membrane protein of Moraxella catarrhalis.
 CC CC The 200 kda outer membrane protein (II) has antibacterial activity and
 CC CC can be used in vaccines. (III), and its truncated versions, are used as
 CC CC immunogenic compositions and vaccines to protect against M. catarrhalis
 CC CC infections, particularly otitis media in humans. (II) is also used as
 CC CC antigen in immunoassays for detecting specific antibodies (Ab), and to
 CC CC generate Ab. (I) are used for recombinant production of (II) and its
 CC CC fragments are used as probes for identifying/cloning 200 kda protein
 CC CC genes from other strains, and for diagnostic detection of M. catarrhalis.
 CC CC (I) makes possible production of large amount of recombinant immunogens.

CC Expression of truncated versions of (II) reduces toxicity of the protein
 CC towards the Escherichia coli host. The present sequence represents the
 CC M. catarrhalis strain 4223 genomic 200kda protein, which is given in the
 CC exemplification of the present invention.
 CC XX
 CC Sequence 2047 AA:
 CC
 CC Query Match 68.9%; Score 7380.5; DB 22; Length 2047;
 CC Best Local Similarity 68.3%; Pred. No. 0;
 CC Matches 1557; Conservative 117; Mismatches 217; Indels 387; Gaps 21;
 CC
 CC 1 MNHIVKIFNKATGTFPAVAEYAKSHSTGGSCATGCVSVPRLSPFARIALAVLIGAT 60
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 1 mnhivkifnkatgtfpa va e y a k s h s t g g s c a t g c v s v p r l s p f a r i a l a v l i g a t 60
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 61 LMSAVA-----GIGISEADGCKGKANARDKSIAIGDIALQALGSOSIAGDKRIYH 112
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 61 lmsava y q q k d t h i a l g e g n p r r s g t a k a d g d r a l a g e n a n a g q a l a g s n k y 120
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 113 NSNNNANIGAKASGNEISIALGIDVLAGHSIAGSDLYL-----KKEYVOQISELL 165
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 121 ngsldkigtatqgesiaigdvkassgdaslaigsddhllqhpnpkpxgtllndi- 179
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 166 PIIRGOKALNDIYQIADTNLOKYYRRTTHAGHASTAVAGAMSYAKGHSNMFGRATAGTY 225
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 180 --lghavllkeirsskndv-kyrrtlasgnastavagamsyagqhfnaftatatsay 236
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 226 SLAVGLATATKAASSIAGVAGSNAOIGFAATVAGSGTQVNLNRGIALFGSQVLQKNDVN 285
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 237 slavglataegqslatigsdatssslgalaigaqraqgslatigqsvvtsqsh--- 293
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 286 AANVR-AYAPDDNOPIDNRKATFKAGATDVESIGNSNGDSIRKRIYVAGASADTDAY 344
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 294 --nsrpaylp-nbqalpbkfgatnntkagpl-----sigsnslikrlinvagavntkdav 345
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 345 NVAGLKAVALA-NROTFKGDSDNNRVEKGLKTLITTGAGTSLTDHNIYVONGD- 402
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 346 nvaqlkaavlwakerrltfsgddnstavkigldntlikgsaetnaalnmivvkeadn 405
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 403 -GLKQLAETLTSIKMVTENTLPAKEKVTVGK----- 433
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 406 sglkvklaktlnltvnttllaattvkysessstaelalsaltftgpnstgstskt 465
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 434 -----TRLTDTKIGFTNDMNCIDSKPYLDK-----D 460
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 456 vygvnyvkftnaetaaigttrlttrdkigfardgdvdkqapyldkqjkygsvallid 525
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 461 TGIHAGGOKITKLAVGVDDDAATYQOLK----- 489
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 526 nglagnkkslnakgssandavllqjkaakprlnagaglsyvtpteisvda ksnvtp 585
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 490 -----RVNQTAESALQTFYKVY 507
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 586 tynlgvktelnsdgtstsksvksgtmslivaehlasylneurtadalsafvke- 644
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 503 CKNGNDANDSKITTVGNKPKPDGTQVNTFLKUGENGVDVTEETNGVTFECLNONGKITVG 567
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 645 -----edddanaaltvakdctknagavsilklkyngnltvaclkdytvlfigdsqslfig 700
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 568 NSTLNDGLSVKNNNSMKOYOVGADGITFPDISKRSKAGIEENTTRTRGIGFANNVGS 627
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 701 kstlndglvtvkd- -negdvga ng i k f l n v g s n p g t g j a n a r t l r ----- 747
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 628 LDANKPRLTPTGINAGKEKELTNVOSAINPATNGQLPFMRRLSTANTEKSGSATIKDLY 687
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 748 ----- 747
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 688 NLSQVPLTFAGDTGPNVTKLIGELIKVKGKGTADDLTKNNIGVADSDNSLTVKLAKT 747
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 748 ----- 747
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 748 LSDDAVNTKTLTASDKVTVDSGNNNTAKLQNGDLTFESKQMTGATPATNSKTIQVGLKFT 807

Db 748 ----- 747
 QY 808 DNNGIALDGTYYITKDVAFKAKODSKSPYIDKDKLKVGEVETTTNGINAGKAITG 867
 Db 748 -----dkigtsgsdgavdlnkpyldgdklgygnvklntglnngykaltg 792
 QY 868 LSNLTLDATNATGHTVQLGIVDSTDKTAAASIDVNLNAGNKLKNNGDADDEVSTYVTD 927
 Db 793 lspclpsiaodssrnlielngticdkdksnaasindlntgfnlknmpidftvscydivd 852
 QY 928 FINGNATTAAYTD -GKASKAVADVNDGTTHLTGADGNKNOIGVKTTLTKTDAGD 985
 Db 853 fangnatlatvthdntsksvkyvduvddtlltltgtdnk -klykktcllknksangn 911
 QY 986 KATFVSNVSGDALKLINAOLADNLNLAGEIRTKGTATATOTFOYKVKKENGDDND 1045
 Db 912 latfnvssdedalvlnakdiaenlntlakeihltkgtadaltqftvckkvdennad-d 970
 QY 1046 ADTITVKGDAKTNOVNTLTKLKGKNGLDIOINKDGTVEGINTFOSGLKAGNNTTLNNGLS 1105
 Db 971 analtvgknaangvntlltkgenglnlktokngtvtlftgntsglkaq -ksclndggls 1029
 QY 1106 IKNTAGNEQIQVADGVKFAKV -NNGVVGAGIDGTRIRTRDEIGFAGTNGSLDKSPHLS 1164
 Db 1030 iknptsgseqlvagadvkfakvnmngvvgagldgttrlttrdelgftgngslkskphls 1089
 QY 1165 KDGINAGKKTITNOSGEIIONSDAVTGKITYDKTELENKISSPRAKTONSLHESVA 1224
 Db 1090 kdglmaagkkltnlsgselqnsndavtggklydtklelnklsstakqnaishetva 1149
 QY 1225 DEQGNFTVSNPSSYDTSKTDVITFPAGENGITTKVKNGVVRGIDOTGTLTPKLTVG 1284
 Db 1150 deggnltfvenpysytsktsdvtitfagengitltnkngvvrvgldgtqltprkltvg 1209
 QY 1285 NNNKGIVIDSQNCNTITGLSNTLANVNDKGSVTRTEGKIIKDEKTRAASIVDLS 1344
 Db 1210 nmngkglvldsqngqcltglntlanvndkgsvtrtegnlikdedktraasivdls 1269
 QY 1345 AGFNLGNGEAVDVFSTYDVNFADGNATAKATYTDTSKTSKVYDVYVNDVDTIEKDK 1404
 Db 1270 agfnlsgngeavdvtvscydvntadgnatcakvtyddtsktskvvydvndvdtlievdkk 1329
 QY 1405 KLGKVTTLTSTGTGANKFALSNOATGDLVKAADIVAHNTLSGDIOTAKGASQANSSA 1464
 Db 1330 klgykttllstgtgankfalsngatgdalvksadivahntlsdgtlqakgsgansea 1389
 QY 1465 GYVDADGNKVIYDSTDNKYYOAKNDGTVDKTEVAKDKLVAAQOTPDGTLAQMNVKSVIN 1524
 Db 1390 gyvdadgnkviydstdnkyyqakndgtvdktkvakdklvaagqtpdgltlaqnmvksvln 1449
 QY 1525 KEQVNDANKKOGINEDNAFPGLEKASDNKTKNAAVTVGDNLAAVQOTPLTFGDPDGTGA 1584
 Db 1450 keqvndankkgylnednafvglekaasdnktknaavtvvgdlnaavqotpltfagdpdgta 1509
 QY 1585 KKLGETLITINGOTDTNKLTDNNINIGVAGTGTFLAKLTLNLNSNAGGTIKIDKGVYS 1644
 Db 1510 kklgetlittggtctdnkltdnnlgyvagtgtflvklakltnlnsvnggtctlddkgys 1569
 QY 1645 FVDSGGAKANTPYLASANGDLGKAYISNKGKTKOTDAAANVOOLEVNRNLGLGNAGND 1704
 Db 1570 fvdssggaakanpylasangldlgkayisnvgkgtktdaanvqglnevrnlglgnagnd 1629
 QY 1705 NADGNVNIADIKDKPSSGSSSNRTYKAGTVGCGKGNDEKLANLGVOVYDVKXGNN 1764
 Db 1630 nadgnvniadikdkpssgsssnrtvlykagtvlygkgnvnteklatgglyvgydkgnan 1689
 QY 1765 GDLSNVWVKTQDKSSKALATATYNAAGOTVYLTNNNAEALIDRINEGIRFEHNDGNOEP 1824
 Db 1690 gdlsvnvwktqdksskallatynaagotvlyltnnpaeadrlnegjirfehvndngnep 1749
 QY 1825 VVGGRNGIDSSAGKHSVAIGFOAKADGEAANAIGROFOAGNOSIAIGDNAQATGDOSTA 1884
 Db 1825 vvggrngidssagkhsvaigfoakadgeaanaigrofoagnosiaigdnaqatgdoستا 1884

Db 1750 vvggrngidsaasagkhsvalgfgakadgeaavaigrqtgagngslatgdnagatgdsia 1809
 QY 1885 IGTGNVYVAKHSGAIGDEBETVAKADNSYSVGNNOFTDAIQTVEVGVGNNTITYTESNVAL 1944
 Db 1810 igtgnvvaakhsvalgdpbtvakadnsysvgnnqftdaiqtvtvfgvgnntitytesnval 1869
 QY 1945 GSNLSASCTHAGTOAKKSDGTAGTTTGTAGTGVKGFAGOTPAVGAASVGSAGARRION 2004
 Db 1870 gsnlsascthagtkakksdgtagttttagatgvtkygfgatvavavgsaagarrion 1929
 QY 2005 VAAGEVSATSDAVNSGSLYKATQGIANAFTNELDRIHQENKANAGISSAMAMASMPQA 2064
 Db 1930 vaagevsatcdavngsglykatgslanaetnelhrihqenkanagissamamasmpqa 1989
 QY 2065 YTPGRSMVTGGIATNHCQAVAVAGLSKSDNCGWFKINGASADTPOGHVGAAGAFHF 2122
 Db 1990 ytpgrsmvtglatnngqavavaglskdsngqwfkinsadtlghvgaavagfhf 2047
 RESULT 3
 AAB69135
 ID AAB69135 standard; Protein; 2053 AA.
 XX
 AC AAB69135;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE M. catarrhalis strain Q8 200kda protein SEQ ID NO:9.
 XX
 KW Moraxella catarrhalis strain Q8; major outer membrane protein;
 KW 200kda outer membrane protein; antibacterial; immunogenic; infection;
 KW otitis media; detection.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO200107619-A1.
 XX
 PD 01-FEB-2001.
 XX
 PE 26-JUL-2000; 2000WO-CA00870.
 XX
 PR 27-JUL-1999; 99US-0361619.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
 XX
 DR WPI: 2001-159722/16.
 XX
 DR N-PSDB: AAF59104.
 XX
 PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
 XX
 PT useful in protective vaccines and for diagnosis
 XX
 PS Claim 1: Fig 4A-V; 247pp; English.
 XX
 The present invention describes an isolated and purified nucleic acid (I)
 CC that encodes a 200 kda outer membrane protein of Moraxella catarrhalis.
 CC The 200 kda outer membrane protein (II) has antibacterial activity and
 CC can be used in vaccines. (II), and its truncated versions, are used as
 CC immunogenic compositions and vaccines to protect against M. catarrhalis
 CC infections, particularly otitis media in humans. (II) is also used as
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to
 CC generate Ab. (I) are used for recombinant production of (II) and its
 CC fragments are used as probes for identifying/cloning 200 kda protein
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.
 CC (I) makes possible production of large amount of recombinant immunogens.
 CC Expression of truncated versions of (II) reduces toxicity of the protein
 CC towards the Escherichia coli host. The present sequence represents the
 CC exemplification of the present invention.
 CC
 CC Sequence 2053 AA;

Query Match 68.7% Score 7357.5 DB 22 Length 2053;
 Best Local Similarity 70.9% Pred. No. 0;
 Matches 1574; Conservative 117; Mismatches 264; Indels 265; Gaps 35;

QY 1 MNHIVYFNKATGTGMAVAEYAKSHSTGGSCATGQVGVRLTSPARIALAVLYGAT 60
 1 mnhivylfnkatgtgmaaveayakshstggscatgqvgvrltsparialavlygat 60
 DB 1 mnhivylfnkatgtgmaaveayakshstggscatgqvgvrltsparialavlygat 60
 QY 61 LNSAYAGIGISEADG-----KGANAKGDKSIAIGDIAQALGSSOIAIGDKIV 111
 61 lnsayagigigtseadg-----kganakgdksiaigdiaqalgssoiaigdkiv 111
 DB 61 lnsayagigigtseadg-----kganakgdksiaigdiaqalgssoiaigdkiv 111
 QY 112 HNSNNNNIGKAKSGNSIAIGDYLASGASHAISTAGSDDIYKKEFYQOISELLPIIRGO 171
 112 hnsnnnnigkaksgnsiaigdylasgashaistagssdiykkefyqoisehellpiirgo 171
 DB 121 pnsngnvgshakgneshlaigdvlaegdaslaigsddilylr-nldlknethkllhgh 179
 121 pnsngnvgshakgneshlaigdvlaegdaslaigsddilylr-nldlknethkllhgh 179
 QY 172 KALNDIQLADTNLQKRYRTHAQGHASTAVGAMSYAKGHEFNAFGRATAGEYSLAVGL 231
 172 kalndiqladtnlqkryrthaqghastavgamsyakghefnafgratageyslavgl 231
 DB 180 elkkigtstvgk-lkyrrtaaghaslavamsyagghhsnaigtvataeayslavgl 238
 180 elkkigtstvgk-lkyrrtaaghaslavamsyagghhsnaigtvataeayslavgl 238
 QY 232 TATKAASSTAVGSAQALGFATAVGSGTOVNLNGIALGFSQVLOKDNVANAANVRA 291
 232 tatkaasstavgsaqalgfatavgsgtovnlngialgfsqvlokdndvnaananvra 291
 DB 239 aagatkgsslavgsnakanaifaataigntvnlgrvaylfgsqldrdnndas---a 295
 239 aagatkgsslavgsnakanaifaataigntvnlgrvaylfgsqldrdnndas---a 295
 QY 292 YAPDNDQPIQNRKATFKNATDVESIGNSNGND-STRRKTIYNGASADTDANVNAOLK 350
 292 yapdndqpiqnrkatfknatdvessignsngnd-strrktiyngasadtadanvnaolk 350
 DB 296 yvp-lgktdaqykatctvgdstdlfsgnsnmslrrkllnvagssrdtdavnaqkl 354
 296 yvp-lgktdaqykatctvgdstdlfsgnsnmslrrkllnvagssrdtdavnaqkl 354
 QY 351 EAVLANQOITFEKGDSDSN--NRVEKGLGKTLITGAGTSAQLDTHNIGVONGGLEVQL 408
 351 eavlanqoitfekgdsdsn--nrvekglgkltitgagtsaqldthnigvongglevql 408
 DB 355 lveelanrktlffkgdghmsnsverglgnltlkgdgaqthalteanlgvntdgnlkvkl 414
 355 lveelanrktlffkgdghmsnsverglgnltlkgdgaqthalteanlgvntdgnlkvkl 414
 QY 409 AETLSKMTTEN-----LT-----AN 426
 409 aetlskmtten-----lt-----an 426
 DB 415 akeltgltsvatnktivsnntnnnaelqsggltfspitgktdktyvsldglkftndsn 474
 415 akeltgltsvatnktivsnntnnnaelqsggltfspitgktdktyvsldglkftndsn 474
 QY 427 EKVAVGTRRLTDKIGFTMDKNGIDESKRPYDKDTGTHAGGQKTKTLAGVDDDAATYG 486
 427 ekvavgtrrltdkigftmdkngideskrpydkdtgthaggqktktlagvdddaatyg 486
 DB 475 slakgttrrlkklkigtatgndgydeskpyld----- 506
 475 slakgttrrlkklkigtatgndgydeskpyld----- 506
 QY 487 QLRKVNQTAESALQTFVKKVDKNGNDANDSKITTVGKNKRPDQYVNTLKLKGENGV 546
 487 qlrkvnqtaesalqtfvkkvdkngndandskittvgknkrpdyvntlklkengenv 546
 DB 507 ----- 506
 507 ----- 506
 QY 547 TTEINGVTGELNONGLTGNGSTLNDGJSVKNTNSKQIQVAGADITFTDSSKRPG- 605
 547 tteingvtgelnongltgngstlndgjsvknntnskqiqvagaditftdsskrpg- 605
 DB 507 -----nekikvgnstlinsgsltvntlgnkqiyganglkatvannant 552
 507 -----nekikvgnstlinsgsltvntlgnkqiyganglkatvannant 552
 QY 606 AGIENTRITFDGIGFANNTGSLDANKPRL-----TPRTGNGKRELTVVQS 652
 606 agientritfdgigfanntgslbankprl-----tprtgngkreltvvqs 652
 DB 553 satvlatleekligfatgndgydegpylkerlkvgrveitdsginsgnhltgl- 610
 553 satvlatleekligfatgndgydegpylkerlkvgrveitdsginsgnhltgl- 610
 QY 653 AINPATNGQGLDFMNRSLSTANTEKSGSAATIKDLYNLSQVLPTRAGTGCPVATKLEI- 711
 653 ainpatngqgldfmnrslstanteksgsaatikdlynlsqvlptraagtgcgvatklei- 711
 DB 611 -----tng-----iantd-----avtlkqklkad- ptnlaagd- galsinsngdlv 648
 611 -----tng-----iantd-----avtlkqklkad- ptnlaagd- galsinsngdlv 648
 QY 712 -----LKVGKGTATADLTGKNNIGVADSTDSITVLAATLSD-LDAVNTKT 758
 712 -----lkvgkgtatadltgknnigvadstdsitvlaatlslsd-ldavntkt 758
 DB 649 dsagntitpnyisvktklinsngtsgnktfsvanahnglv-takdladylkvne- 705
 649 dsagntitpnyisvktklinsngtsgnktfsvanahnglv-takdladylkvne- 705
 QY 759 LTAASKVYVDSGNNTAKIQNDLTFPSKONTGATPA--TNSKTITGVDGLKFTDNNGIALDG 816
 759 ltaaskvyvdsngnntakiqndltfpskontgatpa--tnsktitgvdglkftdnnbialdg 816
 DB 706 -----tadaalpsfkvqngd-----nsnaalvvgktdngktfnt- lklkengvnl- 750
 706 -----tadaalpsfkvqngd-----nsnaalvvgktdngktfnt- lklkengvnl- 750
 QY 817 TTYITKDVGAQKDGSDSKRPYLDKDLKVGVEYITNFI-----NAGGA 864
 817 ttyitkdvgakdgdsdkrpyldkdlkvgeyitnfi-----nagga 864
 DB 751 ttnratgtvlt-----gldsgng- ltpklvgs- dtngnlviiegypsaadgnstknj 801
 751 ttnratgtvlt-----gldsgng- ltpklvgs- dtngnlviiegypsaadgnstknj 801
 QY 865 ITGLSNTLTDATNATGHTVLTGIVDSTDKTAAISIGDVLNAGFNKKNNGAKDFVSTYD 924
 865 itglslntltdatnattgthvtlgtivdstdktaaisigdvlnagfnknknngakdfvsty 924
 DB 802 lkgslptlpslaspsgrlalnigtltleekdknaasiddvlnagfnlknngkdkdfvsty 861
 802 lkgslptlpslaspsgrlalnigtltleekdknaasiddvlnagfnlknngkdkdfvsty 861

QY 925 TVDFINGNATKATKYVD--GKASKVAYDVNVGDTTITLHGADGNKNOIGVTKTTLTKDA 982
 925 tvdfingnatkakyvd--gkaskvaydvnvgdttitlhgadgnknnoigvtkttlktkda 982
 DB 862 tvdfingnatkakyvdngstksaydvnvdektlftlg-dngklqyglkaktelst 920
 862 tvdfingnatkakyvdngstksaydvnvdektlftlg-dngklqyglkaktelst 920
 QY 983 KGDRAINFSVNSGDDKALINARDIADNLTLAGELRNTKGTADTALQTFQVKKVENGDD 1042
 983 kgdrainfsvnsdgddkalinaradiadnltlagelrntkgtadtalqtfqvkkvengdd 1042
 DB 921 ngn-attfstd--ddhalvksadiagnlntlaeehltkygtantalqitfvkvkden-dk 976
 921 ngn-attfstd--ddhalvksadiagnlntlaeehltkygtantalqitfvkvkden-dk 976
 QY 1043 DNDADITVKGAKTQNVNTLKLKNGSLDIOGNKDGVTPEININOSGKAKGNNTLNN 1102
 1043 dndaditvkgaktqnvntlklkngslidiognkdgvtpenininosgkakgnntlenn 1102
 DB 977 addtnalvvgkqtskvntklkqngldlkdqgvltgintqsglkgadsctllmn 1036
 977 addtnalvvgkqtskvntklkqngldlkdqgvltgintqsglkgadsctllmn 1036
 QY 1103 GLSKRTAGNEQIOVGADGVKFAKVNNGVVGAGIDGTRITRTDEIGFAGTNSLDRSKPH 1162
 1103 glskrtagneqiovgadgvkfvakvnngvvgagidgtritrtdeigfagtnsldrskph 1162
 DB 1037 gsklntaeneqiyvgaqdyk-famvnygvvgagldgttirtdeigfagtngslkskph 1096
 1037 gsklntaeneqiyvgaqdyk-famvnygvvgagldgttirtdeigfagtngslkskph 1096
 QY 1097 lskglnaagkkltniqsgelaknsndavtggklydlktelenklstakqnslhfs 1156
 1097 lskglnaagkkltniqsgelaknsndavtggklydlktelenklstakqnslhfs 1156
 QY 1163 LSKDGINAGKKTITNIGSEIAONSNDNAVTVGKIDYLTLEENKJSTAKTQNSLHFS 1222
 1163 lskdginagkktitnigseiaonsndnavtvvgkidyltleenkjstaktqnslhfs 1222
 DB 1283 VADGQNNFTVSNPYSYDTSKTDVITFAGENGITTKYKGVNVGVGIDQTKLTPTKLT 1282
 1283 vadgqnnftvsnpysydtstktvdtfagengittkykvngvvgvgidqtkltptklt 1282
 QY 1157 vadegnnftvsnpysydtstktvdtfagengittkykvngvvtvgldqtkytltpkl 1216
 1157 vadegnnftvsnpysydtstktvdtfagengittkykvngvvtvgldqtkytltpkl 1216
 DB 1283 VGNNGKGIYIDSONCONITGLSNTLANVTNDKGSVFTTEGKIIDKDKTRAASIVDV 1342
 1283 vgnngkgiyyidsonconitglsntlanvtnndkgsvftteggkiidkdktraasivdv 1342
 QY 1217 vgnngkgyivnsqngntltglsntlanvtnndkgsvrttegnlkkedktraasivdv 1276
 1217 vgnngkgyivnsqngntltglsntlanvtnndkgsvrttegnlkkedktraasivdv 1276
 QY 1343 LSAFNLONGEAVADVSTYDVNFADGNATKATKYVDYDDTSKTYVNVVDFTTIEVK 1402
 1343 lsafnlongeavadvstydvnfnadgnatkakyvdyddtsktyvnnvddfttievk 1402
 DB 1277 lsaqfnglqgeevdfvscvdyvfnangnttktakvyddtsktyvnnvddfttievk 1336
 1277 lsaqfnglqgeevdfvscvdyvfnangnttktakvyddtsktyvnnvddfttievk 1336
 QY 1403 DKRLGKTTTLITGTCANKFALSNOATGDALYKADYVAHLNLTSGDIOTKKASQANS 1462
 1403 dkrlgktttlitgtcankfalsnoatgdalylkadyvahlnltsgdiotkkasqans 1462
 DB 1337 dkrlgktttlitgtcankfalsnoatgdalylkadyvahlnltsgdiotkkasqans 1396
 1337 dkrlgktttlitgtcankfalsnoatgdalylkadyvahlnltsgdiotkkasqans 1396
 QY 1463 SAGYVADGNKVIYDSTDKNVYQAKNDGVYDKTEKVAKIKLVAQAOTPGTLAOMNVKSV 1522
 1463 sagyvadgnkviydstdknvyyqakndgvdydktekvakiklvaaotpgtlaomnvksv 1522
 DB 1367 sagyvadgnkviydstdknvyyqakndgvdydktekvakiklvaaotpgtlaomnvksv 1456
 1367 sagyvadgnkviydstdknvyyqakndgvdydktekvakiklvaaotpgtlaomnvksv 1456
 QY 1523 INKEQVNDANKKOGINEDNAFVKGLEKAASDNKTKNAAVTVGDNLNVAOTPLTFAGDTGT 1582
 1523 inkeqvndankkoginednafvkglekaasdnktknaavtvgdnltnvaotpltfagdtgt 1582
 DB 1457 inkeqvndankkoginednafvkglekaasdnktknaavtvgdnltnvaotpltfagdtgt 1516
 1457 inkeqvndankkoginednafvkglekaasdnktknaavtvgdnltnvaotpltfagdtgt 1516
 QY 1583 TAKKIGETLITRKGGOTDKNLTNNIGVAVGTPGVVLAADLTNTNSNAGGTRTDDKG 1642
 1583 takkigetlitrkggotdknltnnigvavgtpgvvlaadltntnsnaggtrtddkg 1642
 DB 1517 takkigetlitrkggotdknltnnigvavgtpgvvlaadltntnsnaggtrtddkg 1576
 1517 takkigetlitrkggotdknltnnigvavgtpgvvlaadltntnsnaggtrtddkg 1576
 QY 1643 VSPVDSGQAKANTPYLSANGLDLGKVTISNVGKTRKDPDAANVQOLNRYNRLGLGNAG 1702
 1643 vspvdsqgakantpylsangldlgkvtisnvvgktrkdpdaanvqolnryynrlglgnag 1702
 DB 1577 vspvdsqgakantpylsangldlgkvtisnvvgktrkdpdaanvqolnryynrlglgnag 1633
 1577 vspvdsqgakantpylsangldlgkvtisnvvgktrkdpdaanvqolnryynrlglgnag 1633
 QY 1703 NDNAQGNQVNIADIKKDPNSGSSSNRTVIAKAGTVLGGKGNNDTEKLATSGVOYGVDDKDN 1762
 1703 ndnaqgnqvniadikkdpnsgsssnrtvliakagtvlggkgnndteklatsgvoygvddkdn 1762
 DB 1634 ndnaqgnqvniadikkdpnsgsssnrtvliakagtvlggkgnndteklatsgvoygvddkdn 1693
 1634 ndnaqgnqvniadikkdpnsgsssnrtvliakagtvlggkgnndteklatsgvoygvddkdn 1693
 QY 1763 ANGDLSNWWYKTKDQSGKALLATYNAAGOTNLTNNPAPADIRTEOGITREPHVNDQNO 1822
 1763 angdlsnwwytkdqsgkallattynaagotnltnnpapadirteogitrephvndqno 1822
 DB 1694 angdlsnwwytkdqsgkallattynaagotnltnnpapadirteogitrephvndqno 1753
 1694 angdlsnwwytkdqsgkallattynaagotnltnnpapadirteogitrephvndqno 1753
 QY 1823 EPVVOGNGTIDSSASGHSVATGFOAKADGEAVALIGRTOAGNOSTAIGDNOAGTGDOS 1882
 1823 epvvoengtidssasghsvatgfoakadgeavaligrtoagnostaigdnogtgdos 1882
 DB 1754 epvvoengtidssasghsvatgfoakadgeavaligrtoagnostaigdnogtgdos 1813
 1754 epvvoengtidssasghsvatgfoakadgeavaligrtoagnostaigdnogtgdos 1813
 QY 1883 IATGTVNAVAGKHSAGIGDPTVAKADNSYVGNNOFTDATOTDVGVGNNITVYESNSV 1942
 1883 iatgtvnavagkhsagigdpvavakadnsyvgnnofdtatodtvgvgnnitvyesnsv 1942
 DB 1814 iatgtvnavagkhsagigdpvavakadnsyvgnnofdtatodtvgvgnnitvyesnsv 1873
 1814 iatgtvnavagkhsagigdpvavakadnsyvgnnofdtatodtvgvgnnitvyesnsv 1873
 QY 1943 ALGSNSAISAGTHAGTOAKKSDGTACTTTAGATGVKAGAGOTAVGAVSVGASGAEKRI 2002
 1943 algsnsaisagthagtoakksgdtactttagatgvkagagotavgavsvgasgaecri 2002
 DB 1874 algsnsaisagthagtoakksgdtactttagatgvkagagotavgavsvgasgaecri 1933
 1874 algsnsaisagthagtoakksgdtactttagatgvkagagotavgavsvgasgaecri 1933
 QY 2003 QNVAAGEVSATSDAVANGSLYKATOGIANATMELDHR.LHONENKANAGISSAMAMASMP 2062
 2003 qnvaaagevsatsdavnangslkyatogianatmelhr.lhonenkanagissamamasmp 2062

Db 1934 qnvaagevsatsctdavnsgslykatqslanaetndhrhnenkanagissamamamp 1993
 QY 2063 QAYIPGRSMVGTATNHGOGAVANGLSKLSDNGOWFKINGSADTQGHVGAAGAGPHF 2122
 Db 1994 qaylprismvgtatnhgogavavglsklsdngowfwikingsadltghvgaagagfthf 2053

RESULT 4

AAB69133
 ID AAB69133 standard; Protein: 1992 AA.

AC AAB69133;

DE 24-APR-2001 (first entry)

XX M. catarrhalis strain 4223 lambdaDEMBL3 clone 200kDa protein SEQ ID NO:3.

XX Moraxella catarrhalis strain 4223; major outer membrane protein;

KM 200kDa outer membrane protein; antibacterial; immunogenic; infection;

KW otitis media; detection.

XX Moraxella catarrhalis.

XX WO200107619-A1.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA00870.

XX 27-JUL-1999; 9905-0361619.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

DR N-PSDB; AAF59100, AAF59101.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

XX useful in protective vaccines and for diagnosis -

XX Example 3; Fig 2A-W; 247pp; English.

CC The present invention describes an isolated and purified nucleic acid (I) that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis. The 200 kDa outer membrane protein (II) has antibacterial activity and can be used in vaccines. (III), and its truncated versions, are used as immunogenic compositions and vaccines to protect against M. catarrhalis infections, particularly otitis media in humans. (II) is also used as antigen in immunoassays for detecting specific antibodies (Ab), and to generate Ab. (I) are used for recombinant production of (II) and its fragments are used as probes for identifying/cloning 200 kDa protein genes from other strains, and for diagnostic detection of M. catarrhalis. (I) makes possible production of large amount of recombinant immunogens. Expression of truncated versions of (II) reduces toxicity of the protein towards the Escherichia coli host. The present sequence represents the M. catarrhalis strain 4223 lambdaDEMBL3 clone 200kDa protein, which is used in the exemplification of the present invention.

CC XX Sequence 1992 AA;

Query Match 66.4%; Score 7113.5; DB 22; Length 1992;

Best Local Similarity 67.6%; Pred No. 6.8e-296; Mismatches 216; Indels 387; Gaps 21;

Matches 1503; Conservative 117; Mismatches 216; Indels 387; Gaps 21;

QY 56 VIGATLNGSAYA-----GIGISEADGKGGANARCKSIAIGDIAQALQSQTATG 107

Db 1 vlgatlsgeayagkdkthlaigeqqprirgltakadgdratlaigenanagsgalaigs 60

QY 108 NKTIVHNSNNANIGAKASGNESIAIGDVLAGSHASIAIGSDLYL-----KRTTYVO 160

Db 61 snktvngssldkigtatqgeslaigdvkaasdaslaigsodlhlldghnphkphkgtl 120
 QY 161 ISELLPIIRGOKALNDIYQADTNLQKRRTHAOGASTAVAGMAYAKHFSNATRTAT 220
 Db 121 indl--inghavlketrskdndv--kyrrtlasgnastavagmaysagghfhsnalgtrlat 176
 QY 221 AEGTYSIAVGLTITAKAASSIAVGSNAQAIQFAATVAGSGSTQVNLNRGIALGSGSVLQK 280
 Db 177 aksayslaavglaaetlaeqgsilaigsdatsessilaigatqatqaglaigqsvvsgs 236
 QY 281 DNDVNAINVR-AVAPDNDQPIIDNRKATFEKNGATDVFSGNSGNDISIRKTIINWAGSA 339
 Db 237 dn-----nsrpaytp-ntgaldpkfgatuntkagpl-----sigsnsikrklinvagayn 285
 QY 340 DTDVAVNAQLEAVRIA--NQITFEKGDSSNNRVEKGLGTLTITGCAQTSLTDHIGVY 398
 Db 286 ktdavnyaqgleavkvkwaekerrlftgqddnsdvkigldnllikggaetnaltdmniygv 345
 QY 399 QNGD--GLKQLAETLTLKLVNTENTANEKVTYVK----- 433
 Db 346 kdadngslkvklakltnlnlevntltlnaltlvkvgssstlaellsdslftqpnsgq 405
 QY 434 -----TLLTPDKIGFTMDMNGIDESKPYLDK----- 459
 Db 406 stktvygvnvgvftmaettaigttrtrdkigtaragdvekeqpyldkqklvgsv 465
 QY 460 ---DTGIIHAGGQKRTKLVAGVDDDAATYGLK----- 489
 Db 466 altidngidagunkklslnakgssandevltleqlkaekprlnaagajsvirpelsvskag 525
 QY 490 -----KVNQRTESALQTF 502
 Db 526 nvtaipylnigvktelnsdgtscdfsvksgslmsjlvtaehlsylnvnrtdsalsgaf 585
 QY 503 TVKKVDKNGDANDSKITITVGNKNNKPDGTQVNTLKLKGENVDTTETNGTFTGNNQNN 562
 Db 586 tvke-----edddanaaitvakotdknagavslklkgknglvackkdgvtvfglsqgs 640
 QY 563 GLTVGNSLNLNDGLSVKNTNSNKOIOVGADGITFTDISNSKPGAGLENTTRTRDGI 622
 Db 641 glitgkstilmdgltvtvk--neqiygavangikftlvngsnpqlgtlanlarlfr----- 692
 QY 623 NNTGSLDANKRPLPTIGINMGKELTVQSAINPATNGQLDPMNRNSTANTEKSSAAT 662
 Db 693 ----- 692
 QY 663 IKDLYNLSQVPLTFPAGDTGPNVTAKLGEILKVGKTTADDLTKNNIGVAVDSTNSLT 742
 Db 693 ----- 692
 QY 743 KLAATLSDLAVENTKTLTASDKVTVDGNNATKALQNGDLTFPSKONTGATPATNSKIT 802
 Db 693 ----- 692
 QY 603 GLKFTDNNGLADGTYTTRKKGVEAKQDGLSKRPLDKKLKVGVEVETTINGINAG 862
 Db 693 -----dkigfagsdgavdtnkpyldgdklgyvnnvltntgtlnag 722
 QY 863 KATGLSNLTLDATNATTTQGIYDSDTKTFRASIGDVLANAGFNLKNNDAKDFVST 922
 Db 733 kaigtlsplpsladsgrnlelgnltgdkksnaasindlntgfnlknmpidfvst 792
 QY 923 YDVFDFINGNATYAKVYD--GKASKVAYDVNVDTGTHLTLGADGKNQNGVCTTTLTKT 980
 Db 852 saangtahnlnvssdeodalvnaekdienltakelhtkqgtadtaqlctfvkvdenn 911
 QY 961 DAKGDKAINEVNSGDDKALINAKDIADNLTLAGELIRNRGTADPRLQTFQVYKVE 1040
 Db 1041 DDDNDADTITVGDAKTQVNTLKLKGNGLDITGNRGTVEGTFINTQSGIKAGNNTTN 1100
 QY 912 nad--danaitvgqkannqvnltlkgengnlnktkdngrvtgintsglkgk--kstln 969

Db 237 dn-----nsrpaycp-ntqaldpkrfgatnmtkagpl-----sigsmskklkinnvgavna 285
 QY 340 DTDVAVNAOLKEAVRLA-NNOITFKGDDSNRRVEKGLGKLTITTGAGOTASLDHNGV 338
 Db 286 ktdavnaqgleavvkwakeritltfgdndstdvkigldicltkysgaetnaldnigv 345
 QY 399 QNGD--GLKQJLAETTLKMTTENTLANEKVYVK----- 433
 Db 346 keadnsglkvklaktlnlnleventtlnalttkvvgssstaeallsdsiltfcpntbsq 405
 QY 434 -----RLTPTDKIGFTNDMNGIDESKPYLDK----- 459
 Db 406 stsktygvngvfkfinaettaigttrirckfigfardgdvdekqpydkkqklykvsv 465
 QY 460 ----DTGIHAGGOKITKTLTGVVDDDAATYGOLK----- 489
 Db 466 ailtidngidagknkkslnlakysandavtleqilkaakprlnaagagisvprtleisvdksg 525
 QY 490 -----KVNOZAESALOTF 502
 Db 526 nvtaptyngvkttelnsdgtaskfayksgqtnslvteahlasylnevnrtadsalgt 585
 QY 503 TVKKVVKNGNDANDSKITTYGKNNKPDGTOYNTLKLKGENGVDTYETNGVTFTGNNQNN 562
 Db 586 tvke-----edddanaltvakdtknagavslkiklygnlvtackdgtvftvglsqds 640
 QY 563 GLTVNGSTLNNDGLSVKNTNSKIOVAGDITFTDISNKGAGIENTRITRDIGRA 622
 Db 641 gltfgstlmdgltykdc--neqiygagngkftfnvngsnpgtylanarlttr----- 692
 QY 623 NNTGSLDANKPRLPTGINAGKELTNVQSAINPATNGQDPMRLSTANTEKSSAAT 682
 Db 693 ----- 692
 QY 683 IKDLNLSQVPLTFAGDTGPNVTKKIGELIKVAGKTTADDLTKNNIGVADSTDSNLTIV 742
 Db 693 ----- 692
 QY 743 KLAKTISLDLAVNTKTLTASDKVTVDSGNNTAKLQNGDLTFESQONTGATPATNSKITGV 802
 Db 693 ----- 692
 QY 803 GLKFTDNNGLADGTYITRKDVYGAFCODSGLDKSPYLDKCLKVGEVEITNGINAG 862
 Db 693 -----dkigfagsdgvadtkpkyldqklqygnvklntngnag 732
 QY 863 KATGLSNTLTDAATNATGHWYQLGIVDSTDKTRASIGDYVNAAGPNLKNNDADPFYST 922
 Db 733 kaltglisptlpslaqssnlelgnltqdkksnaasindlntgfnknmpidifvst 792
 QY 923 YDTVDFINGNATTAKTYVD--GKASKVAYDVWVDTTILHTGADGKKNQIGVKTTLTFT 980
 Db 793 ydtvdfangnattactvthtbanktskvyvdyvndtclltlgtddnk-klyvktcltlnkt 851
 QY 981 DAKGKATIFSVNSGDKALINAKDIADNLNLAGEIRNTKGTADTALOTFQVKKVYKENG 1040
 Db 852 sanngatcltnfnvnsdedalvnaakdaenlnl lakeinhctktadtaqltftvckvden 911
 QY 1041 DDNDNDADTITVVGKDAKTNOVNTLKLKGNGLDIOTNKGSTVFGINTOSGLKAGNNTILN 1100
 Db 912 nad--danaitvgkannnqvntclikgenglnlckdngtvgfintstgkag-ksctln 969
 QY 1101 NNGLSIKNTAGNEOIQVADGVYFAKV--NGVYVAGIDGTTRITRDEIGFAGTNGSLDKS 1159
 Db 970 dgglsiknptgseqiqvgadgkfkavmngvyvgagidgtttritdeigftgngslks 1029
 QY 1160 KPHLSKDGINAGGKKTITNIOSEIQAQNSNDAYVGKITYDLKTELENKISSTAKTAQNSIH 1219
 Db 1030 kphlskdginaagkkltnlqsgelqasndavtggklydlkteleenklststaktagsnsh 1089
 QY 1220 EFSVADGCGNNTFVSNPYSSYDTSPTVITFAGENGITTVKKNKVAVNGIDOTGLTTP 1279
 Db 1220 EFSVADGCGNNTFVSNPYSSYDTSPTVITFAGENGITTVKKNKVAVNGIDOTGLTTP 1279

Db 1090 efsvadeggnftvsnpyssydstsktsdvitfagenglttkvknkvrvvgldgtkgltp 1149
 QY 1280 KLTGVNNGNGGLVIDSONGONTITGLSNTLANVNTDKGVSRTTEGOKITKEDEKTRASI 1339
 Db 1150 klvngnngkglividsngnntltglsntlanvntdnkgsvrttegnlkkedektraasi 1209
 QY 1340 VDVLSAGFNLQNGEAVDVSTVDTVNFADGNAATYAKTYTIDTSTKSVYVDVNDVDTTI 1399
 Db 1210 vdvlsagfnlqngaevdvsvtydvtvntadgnattaktytdtsktskvvydvndvdtlti 1269
 QY 1400 EVKDKKLGVKTTTLTSTGANKFALSNOATGDALVKKASDIYAHNLNTSGDIOTAKGASQ 1459
 Db 1270 evkdkklgvtltltstgankfalsnagtdalvksasdiyahnlntsgdiotakgasq 1329
 QY 1460 ANSSAGYVDADGNKYIYDSTDNKKYVQAKNDGTVDKTEKVAADKILVAQAOTPDGTLAOMNV 1519
 Db 1330 anssagvydadgnkylvdstdnkkygkndgtvdktkavakdklvaqgcpdgtlaqnmv 1389
 QY 1520 KSVINKBOVDANKKOGINEDNAFVKGLEKASDMKTKMAAVTVGDILMAVAQOTPLTFAGD 1579
 Db 1350 ksvinkevndankkqglinednalfvglekaasdnktnaaevlvgdlnavaqcpdltlagd 1449
 QY 1560 TGTTRAKKLGELTLTIGGQDTNKLTDNNIGVAGTDTFVKLAKDLTNLSVNNAGTRID 1639
 Db 1450 tgttrakklgeltltlgqgdtlnkltdnnlgyvagtgdftvklakdltnlnsvnaggtklid 1509
 QY 1640 DKGVAFVSSGAKRNTPYLSANGDLGKXYISNVGKGTCTDAAVQOOLNEVRNLGLG 1699
 Db 1510 dkgvafvssgakarntpylsangdlgkxyisnvgkgtctdaavqoolnevrnlglg 1569
 QY 1700 NAGNDNAGNOVNIADIKKDPSSGSSNRVYIKAGTVYGVGGNNDTEKLANGQVGVNDK 1759
 Db 1570 nagnndnagnovnialdkkdpssgssnrvtlykagtlvlygqymndteklacvgtlqvavdk 1629
 QY 1760 DGNANGDLSNVWVKTOQDKSGRKALLATYNAAGQTNVLTNNPAAEIDRINEGIRFFHVND 1819
 Db 1630 dganangdlsnvwvktqkdksgskallacynaagqtnvltlnpaaeidrlnegriffhvnd 1689
 QY 1820 GNOEPVYVGRNGIDSSAGKHSVAIGPOAKADGEAAVAIGQTOAGQNSIAGDMAATG 1879
 Db 1690 gnoepvyvgrngidssagkhsvaligpoakadgeaavagrtqagngslagdnagatg 1749
 QY 1880 DQSLAIGTGNVAVAGKHSAGIADPSTVKADNSYSGVNNNOFIDATQOTVFGVNNITVTES 1939
 Db 1750 dgsilaigtgnvavagkhsaigdpstvkadnsysgvnnmqfidaqtvtvfygmnlvttes 1809
 QY 1940 NSVALGNSAISAGTHAGTOAKKSDGTAGTTTATGATGVYKGFAGQTAAGVAVSGASGAE 1999
 Db 1810 nsvalgnsaisagthagtoakksdgtagttttatgatgvkykfgfagtaavsgasgae 1869
 QY 2000 RRIONVAAGEVSATSTAVNSOULYKATQGANNTNLDHRIRHONENKANAGISSAMAMA 2059
 Db 1870 rriionvaagevsatstavnsoulykatqganntnldhrirhonenkanagissamama 1929
 QY 2060 SMPQAIYIPGRSMVTGIAITHNGQAVAVAGLSKISDNQCMFKNIGSADTQGHGAAGVAG 2119
 Db 1930 smpqaiyipgrsmvtgialthngqavavaglskisdncmfknigsadtqghgaagvavag 1989
 QY 2120 FHF 2122
 Db 1990 fhf 1992
 RESULT 6
 AAB69137
 ID AAB69137 standard; Protein; 1992 AA.
 AC AAB69137;
 XX
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE M. catarrhalis M56 200kDa protein in PKS348 SEQ ID NO:13.
 XX

| Query Match | Best Local Similarity | Score | DB | Length |
|--|-----------------------|---------|---------|--------|
| Matches 1502; Conservative 118; Mismatches 215; Indels 387; Gaps 21. | 66.4%; | 7110.5; | DB: 22; | 1992; |
| 56 VIGATLNGSAYV-----GIGISPADGKGAGNARCKSTAIQDIAQALQSQTIAIGD 107 | : | | | |
| 1 mlgatlsqgaayqgkdkdtkhtalgegnqprrrstakadgdrralalgaigenaagsgqatalys 60 | : | | | |
| 108 NKIVHNSNNANIGAKAKSGNESIAIGCDVLASGHSIAIGSDLLV-----KKETVQO 160 | | | | |
| 61 snktvngssldkigtatdqtgsaiaigdgkkaasgaasialgsdhlhldghnphkpkgtl 120 | : | | | |
| 161 ISKILPITRGKALNDIYOLADTNLOKRYRTTAOGHASTANVAGAMSTAKHFNPAETRTT 220 | : | | | |
| 121 indl---lghnavlkeirskdndv-kyrtrtasghastavayameyagghfnatqtrrt 176 | : | | | |
| 221 AEGTYSILAVGLTATATAKASSIANGVSNAAQIGFAATVAGSGSTOVNLNRGIALTGFSQVLQK 280 | : | | | |
| 177 aksagylavglataetagsctialgsatssalsgaialgdyttraiqgslalqgsvvtgs 236 | : | | | |
| 281 DNDVNAANVR-AIAPDNDOPIDNRKYATRTKNGATDVFSIGNSGNDSTIRKLIINVGAGSA 339 | : | | | |
| 237 dn-----nsrpaytp-ntaladpkfqtatnltkxgpl-----sigsnsikrkliinvagyn 285 | : | | | |
| 340 DTPDAVVAALKEAVRLA-NRQITFEKGDGSDNNRVEKGLGKRTITTTGGAQTSALTDNHNIGVY 398 | : | | | |
| 286 ktlaavvaqgleavvkwakertrlltfggdndstadvkigldnlltklqgaetnaltdmniyvv 345 | : | | | |

| | | | |
|----|------|--|------|
| QY | 369 | QNGD--GLKYOAEFFLRLSKMTTEMLTANKEVYWK----- | 433 |
| Db | 346 | keadnsglkvklaktlnlnllevntlttaactlvksssstaelisolslftgpnqsg | 405 |
| QY | 424 | -----TRLVTPDKIGFTTDMNGIDESKPYLDK----- | 459 |
| Db | 406 | stsktvygvngvfktnmaetaiaigtlttrllrkijgfaldgdvdekqpyldkqjlkxgsv | 465 |
| QY | 460 | -----DTGHHAGGOKIRKLTAGVVDDAATVGGOL----- | 489 |
| Db | 466 | ahtldngidagunkhslmagssandavtleqlkaakpnlmagaglsavtpealsvdaksg | 525 |
| QY | 490 | -----KVNQFASALQTF----- | 502 |
| Db | 526 | nvtaptylniykctelnsdgtscdfsvkysgjnmslvtaehasylnvemtalsalqsf | 565 |
| QY | 503 | TVKRVKDGNDANDSKLIIRYGRKNKEDGQVWMTLKGKNGVDVTEFNGVTFGLQNN | 562 |
| Db | 566 | lvke-----edddanaaitvakctknagavslklkkgnglvtaclkkgdgvctlglsqds | 640 |
| QY | 563 | GLTVGNSTLNNDGSLVKNKTNNSKOIOVGADGTFITFDISNKPACGIENTTIRTRDGIGA | 622 |
| Db | 641 | glitgksetlmdgtlvctd--neqiygvanqjklftvnvsnpygtjantarltr----- | 692 |
| QY | 623 | NNTGSLDANKRRLTPGTGINAGKELINVOASIPATNGOLDPMNRLESTANTEKSGSAT | 682 |
| Db | 693 | ----- | 692 |
| QY | 683 | IKDLYNLQVPLTEPAGDPGPNVTKLGEILKVGKGTADTLTKNNIGVADSTDNLSLV | 742 |
| Db | 693 | ----- | 692 |
| QY | 743 | KLAKTLSDLDAVNTKTLTASDKVTVDSGNNTAKLQNGDLTFSKONTGATPATNSKTIQVD | 802 |
| Db | 693 | ----- | 692 |
| QY | 863 | GLKFTPDNNGIALDGTYYTITNDKVPFAKQOSLSDKSPRYLDKRLKLVGEVEITTNNGAG | 862 |
| Db | 693 | -----dklqfsgdvavtlnpnyldqklqygnvklntnginaagg | 732 |
| QY | 863 | KAITGLSTMTLIDATNATTGHTQGLGIVDSYDVKTRAASIDVYNAGFNLKNNGDADKPFVST | 922 |
| Db | 733 | kaitglisplpslidsqssrniegntclqddkksnaaslnldlntgfnlknmpdlfst | 792 |
| QY | 923 | YDTPDFINGNATTAFAKYTD--GKASKVAYDVNVDSFTTHLTGAQGNKOIGVKTITLFT | 980 |
| Db | 793 | ydvdfangnatltvthdbanktskvygvnvdctllnlgcdnk--klygkttklakt | 851 |
| QY | 981 | DAKGRKALNFYSVNSGDDKALINKADINDINLTLAGIRMTKRGAPATALOTPOVKVYKENG | 1040 |
| Db | 852 | sangnatlnfivnsdcdalvnaakdaenlntlakeihctkgpadalqfvtvkkvndem | 911 |
| QY | 1041 | DDDDNADRTVYGKAKTQVWVTEKLKGKNDLIDOTKNKGDVTFGINTOSGLKAGNNLTN | 1100 |
| Db | 912 | aad--danaalvygqknamqvnclclxgenglnlktkdngtvlfqntstglskag--kseln | 969 |
| QY | 1101 | NNGLSIKNTAGNEOIQVGADGVKFAKY--NNGVYGAGIDGTTIRTRDELGEFAGTNGSLDKS | 1159 |
| Db | 970 | 3ggjstsknptgseqyqvgaadvkftakvnnngvyvagydgftrttrdelgftgtnslaks | 1029 |
| QY | 1160 | KPHLSKDGIDINAGGKRTTINIGSGELIAONSNDNAVYGGKRYLDKLTLEKKISSTAKTQNSLH | 1219 |
| Db | 1030 | kphlskdgidnagkkltnlqsgelagushdavtgyklydktlelnklsstaktqnslh | 1089 |
| QY | 1220 | ESSVADDEGNNFTFVNPYSYSDTSEKTSADVITPEAGENITFKVVKGVAVRGIDQTFGLTTP | 1279 |
| Db | 1090 | efsvaddegnftvsnpyssjdsteksdvltfgengnlttknkvyyrvydgctgltcp | 1149 |
| QY | 1280 | KLTGVGNNGKGIVIDISQNGQWTTIGLSNLTLANWNTNDKGSVRYTTEOCKIKDEDKTRAASI | 1339 |
| Db | 1150 | kltvgngngkglvidisqngntclglntlanlvndksgsvrttegnlklkdekttraasi | 1209 |
| QY | 1340 | VDVLSAGFVLQNGEAVDPFVSTYDTVNFADGNMTAKVYTDOTSKTSKVYDVVNDDTTI | 1399 |

```

Db 1210 vdlvsagfhlqngseavdvstydvtvfnfagdnattkavdyddtsktskvvyvnnvddtci 1269
OY 1400 EYVDDKLLGKTTTTLSTGTGANKFALSNQATGDALVKAASDIYAHNTLSGDIOTAKGASQ 1459
Db 1270 evkdkklyvktcttctscgcnkfalsnqatgdalvkaasdiyahntlsqdiqtakgaag 1329
OY 1460 ANSSAGYVDADGNKYIYDSTDNKYYOAKNDGVDTKEVAKRKLVAQAQPTPGTTAQMNV 1519
Db 1330 anssagvydadgnkvlydstdnkyyakndgvtckevakrklvaagqptpgttlaqmny 1389
OY 1520 KSYINKBQVNDANKKKGIEDNAFYKGLEKASDNKTKAAATVGDLANVAQPTLTFACD 1579
Db 1390 ksyinkbqvndankkgyiednafykglekaasdnktknaatvgdlnavagptltfagcd 1449
OY 1580 TGTAKKLTETLTIKGQDTNKLTDNNITGVAVAGTGFVKLAKDLTNSNVNAGGTRID 1639
Db 1450 tgtakklteetltikgqgdtcnkltdnnitgvavagtdgfvlakdltnsnvnaggtklid 1509
OY 1640 DKGVSEFVDSGQAKANTPVLASNGDLGKVISNVGKGTDPDAAVQDLNEVRNLGLG 1699
Db 1510 dkgsfvdsgqakanpvlansngdlgkvisnvgkgtddpdaavqdlnevrnlglg 1569
OY 1700 NAGNMDADNQNINADIKKDPMSGSSNRTVYKAGTVLGCKGNNTTEKLTGCVGVYDK 1759
Db 1570 nagnmdadnqyniadikkdpmsgssnrtvylkagtlvgkgnntteklatgigvydk 1629
OY 1760 DGNANGLDSNVWVTKQDKSKALLATYNAAGOTVLTNNPAEALIDRIINEGIRFEHYND 1819
Db 1630 dgnangdlsnvwvktqdkgskallaltynaagotvltnppeaaldrinegirtfhynd 1689
OY 1820 GNOEPVVGKNGSIDSSASGKSHVAIFQAKADGEAAVAIGRQTOAGNOSTAIGDNAQATG 1879
Db 1690 gnoepvvgkngsidssasgkshvalfgakadgeaavaigrqtogngsstaigdnaqatg 1749
OY 1880 DOSIATGNGNVVAGKHSGLIGPSTVYKADNSYVGNNOPTPATGTDFEGVGNNTTVES 1939
Db 1750 dosiatgngnvvagkhsaglpgstvkadnsyvgnnopftpatgtdfegvgnnttves 1809
OY 1940 NSVALGSNSAISAGTHAGTQAKRSKSDGTATTTAGATGTVKGFAGQTFVAVGASVGSABE 1999
Db 1810 nsvalgsnsaisagthagtgakrsksdgtatttagatgtvkgfagqvavgsasvgsab 1869
OY 2000 RRIQNVAAEGVSAISTDAVNSQLYKATQAGANATNEIDHRIHONENKRNAGISSAMAMA 2059
Db 1870 rriqnvaaegvsaistdavnsqlykatqaganaetneidhrihonenkanagissamama 1929
OY 2060 SMPQAYIPGRSMWGTGIATHNGOGAVAVGLSKLSDNGQWFKINSADTQOGVGAAGAG 2119
Db 1930 smpgayipgrsmwgtgiathngogavavglksldngqwklingssadtqogvgaagag 1989
OY 2120 FHF 2122
Db 1990 fhf 1992

```

RESULT 7

AAB69136 standard: Protein: 2314 AA.

AAB69136:

24-APR-2001 (first entry)

M. catarrhalis les1 200kDa protein SEQ ID NO:11.

Moraxella catarrhalis strain Q8: major outer membrane protein;

200kDa outer membrane protein; antibacterial; immunogenic; infection;

Moraxella catarrhalis.

Moraxella catarrhalis.

```

XX 01-FEB-2001.
PD 26-JUL-2000; 2000MO-CA00870.
XX 27-JUL-1999; 99US-0361619.
XX (CONN) CONNAUGHT LAB LTD.
XX Logjampro SM, Sasaki K, Yang Y, Klein MH.
XX WPI; 2001-159722/16.
DR N-PSDB; AAF59105.
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
PT useful in protective vaccines and for diagnosis
PS Claim 1; Fig 5A-Y; 247P; English.
XX
CC The present invention describes an isolated and purified nucleic acid (I)
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
CC The 200 kDa outer membrane protein (II) has antibacterial activity and
CC can be used in vaccines. (II), and its truncated versions, are used as
CC immunogenic compositions and vaccines to protect against M. catarrhalis
CC infections, particularly otitis media in humans. (II) is also used as
CC antigen in immunoassays for detecting specific antibodies (Ab), and to
CC generate Ab. (I) are used for recombinant production of (II) and its
CC fragments are used as probes for identifying/cloning 200 kDa protein
CC genes from other strains, and for diagnostic detection of M. catarrhalis.
CC (I) makes possible production of large amount of recombinant immunogens.
CC Expression of truncated versions of (II) reduces toxicity of the protein
CC towards the Escherichia coli host. The present sequence represents the
CC M. catarrhalis les1 200kDa protein, which is given in the exemplification
CC of the present invention.
XX
SQ Sequence 2314 AA.

```

Query Match 61.6%; Score 6600; DB 22; Length 2314;

Best Local Similarity 57.1%; Pred. No. 7; le-274;

Matches 1460; Conservative 133; Mismatches 278; Indels 682; Gaps 39;

```

OY 1 MNIHYKVFNKATGTENAAVAEYAKSHSTGGSCATGCGSVRTLSFARIALAVLIGAT 60
Db 1 mnhlykvfnkatgttnaavaeyakshstggscatgcvsvrtlsfarialavlignat 60
OY 61 LNSGAVYAGIGISEMDGKCG--GAMARGDKSIAIGDIAQALGQSTAIAGDNK-IVNNSNN 116
Db 61 lnsavayagigisemdkgc---gamargdksiaigdiaqalgsqstaiagdnkivnnsnn 116
OY 117 NANIGAKASGNESIAIGDVLASGSHASIAIGSDLYLKRTTYOQ--ISELPIIRIGOKA 173
Db 117 nanigakasnesiaigdvlasgshasiaigsdlylkrttyoq--iselpiirigoka 173
OY 121 nqkgsnshakkesialsgdvlaegdasialsgsdlyldrnstskypngllstlignhev 180
Db 121 nqkgsnshakkesialsgdvlaegdasialsgsdlyldrnstskypngllstlignhev 180
OY 174 LNDIYQLADPN-LQKVRRTAAGHASTAVGAMSYAKGHSNAGFTRARAEGLYSLAVLT 232
Db 174 lndiyqladpn-lqkvrtrtaaghasavagamsyakghsnagftraraeglyslavlt 232
OY 181 lr---glrdsngsqkyrrlaeghaslavgamayakghafnaelgrtsteagmslavlt 237
Db 181 lr---glrdsngsqkyrrlaeghaslavgamayakghafnaelgrtsteagmslavlt 237
OY 233 ATKKAASSIAVGSNAOALGFAATAVAGSTOVNLRGALGFGSOVLOKXNDVAAAVRXY 292
Db 233 atkkaassiavgdnaoalgfatavagstovnlrgalgfgsovlokdndvaaaavrxy 292
OY 238 bkaekyrtlaigsnagaingalaigadtrvldyalaigysqll---nnnnnnnnkay 294
Db 238 bkaekyrtlaigsnagaingalaigadtrvldyalaigysqll---nnnnnnnnkay 294
OY 293 APDNDOPIDNRKATFENGATDFVSGNSGNDISIRKTIINVGASADPDAAVVAOLKEA 352
Db 293 apdndopidnrkatfengatdfvsgnsngndisirktiinvgasadpdaaavvaolkea 352
OY 295 vpegngsnlkskat-gng---lfsigs-----tlkrklinyaggyedcdavnavglkav 346
Db 295 vpegngsnlkskat-gng---lfsigs-----tlkrklinyaggyedcdavnavglkav 346
OY 353 VRLANRQITFKGDSNNRNVKGLKTLTTGG-AQTSALTD-HNIGV-ONQDGLKVLQA 409
Db 353 vrlanrqitfkgsdnrvkglkltttgg-aqtsaltd-hnigv-onqdgkvlkva 409
OY 347 enlakrgitfkqdgngltvkkigetltikgetgdkltdnnlgyvcladnltgkvlka 406
Db 347 enlakrgitfkqdgngltvkkigetltikgetgdkltdnnlgyvcladnltgkvlka 406
OY 410 ETLTSLKMTTENTLANKETVTK----- 433
Db 410 etltslkmttntlnkxetvtk----- 433
OY 407 knlsgletvstknltasekvtyvgngntaelsgglltftptnascdkcvygtgdlkftd 466
Db 407 knlsgletvstknltasekvtyvgngntaelsgglltftptnascdkcvygtgdlkftd 466

```


| | | | |
|----|------|---|------|
| OY | 434 | -----TFLTUDKIGFNNDMNGIDESPVLXDKPTGHIAGQCKTKTLTAIVDDDAAT | 484 |
| Dd | 467 | nsntaledttrilckdkgisfnkaglyvdenkpyldq----- | 502 |
| OY | 485 | YGLKVKVQGTAE\$ALQTFYVKKVDKNGDANDSKITTVGKNKPDGTQVNTLKLGENGV | 544 |
| Dd | 503 | -----klk----- | 505 |
| OY | 545 | DVTEETNGTVEFLGNQNNGLTVGNSTLNNDSLKYNT--N\$NKQIOVAGDITTFDISNS | 602 |
| Dd | 506 | -----vgnstlmggltvnmntlgssnkqqlvgdgdyklfadvnvn | 544 |
| OY | 603 | KPGAGIEHTTITIDGIGFANNGTSGLDANKPRL-----TPGJNAGGKELTN | 649 |
| Dd | 545 | vsnaakfgttrileeeijfadagdkvdkkspyldkqqlgvgyvkltkdsginnagqaksn | 604 |
| OY | 650 | VOSA-----INPATNGQLDPMNLTSTRANEKSG\$AATIKDLYNLSQV-- | 693 |
| Dd | 605 | vkdatddvadykqlkvvgqdaagalsfrt-----deyqgefislmysnqncpt | 658 |
| OY | 694 | ---LTFAGDGPNNYTKLGEILVKVGKTTADDTLNNIGVAD\$TNSIVTAKLAKTISD | 750 |
| Dd | 659 | fettifagengisl-----sndiakgkvv----- | 684 |
| OY | 751 | LDAVNTKTLTASDKVYVDSGNNTAKLQNGDLTF\$KONTGATPATNSKITGVDLKFTDNN | 810 |
| Dd | 685 | ldpin-----gltcpkl-----lvgsd----- | 701 |
| OY | 811 | GIALDGTPTYIKDKVGFARQKDDGSLDSK\$PYLDDKDKLKVGEVELTNGJNAG\$KALTIG\$LN | 870 |
| Dd | 702 | ---kdgkqlyievga\$ngdtknl-----trglsp | 728 |
| OY | 871 | TJLTATNATGHWYOLG--IVDSTDKT\$RASIGDYV\$NAGFNLKNN\$D\$KAD\$F\$TYDYDF | 928 |
| Dd | 729 | tlrsltnaggyrttegnntltedsedkskaasjgdltngfhlkmsnsvgfrstynvdf | 788 |
| OY | 929 | INGNATTA\$KAYTD--GKASVAV\$AVNDVGTTHL\$TGAD\$KNKQIOGVKTTTLTKTDAK\$DK | 986 |
| Dd | 789 | ldgnatktavyde\$ngtskvydvnmvdektlaeltgdngktrkikgykttltnang-k | 847 |
| OY | 987 | AINSVNSGDDKALIN\$KODLADMLN\$LAGIR\$TKTATPA\$LO\$TQVYKVK\$ENGDD\$D\$DA | 1046 |
| Dd | 848 | atn\$stt--dndalvnakdaenlntlakeihltktdatlaqtflvk-----dgatdd | 900 |
| OY | 1047 | DTIVG\$KDA\$KTN--QVWTLTLK\$KNGJLDTQNTKNDGVFVG\$IMQ\$SLK\$GNTT\$LN\$NNGL | 1104 |
| Dd | 901 | etlcvg\$dg\$ngk\$ntvntllkik\$engjlyva\$cm\$dv\$vtl\$gtlnt\$gslkag\$dstl\$nk\$gl | 960 |
| OY | 1105 | SIRKTAGNEQIOVAGD\$V\$K\$F\$K\$AVNNGV\$VAG\$IG\$D\$G\$T\$R\$IR\$DE\$IG\$AG\$T\$NG\$S\$LD\$K\$P\$H\$S | 1164 |
| Dd | 961 | sikp\$asneq\$vg\$dg\$dyk\$tk\$dvk\$dn\$st\$gld\$g\$tr\$ic\$kd\$g\$lf\$fg\$ng\$sl\$dt\$tk\$phlt | 1020 |
| OY | 1165 | KD-----GJNAG\$K\$K\$T\$T\$IO\$G\$E\$IA\$ONS\$DAV\$G\$K\$IDY\$DK\$TELE\$NK\$IS\$T\$AK | 1212 |
| Dd | 1021 | kdkl\$kvgeveltneglna\$g\$kklt\$ng\$sdlt\$ng\$ndav\$tr\$gv\$dlk\$teles\$ks\$nsa\$ak | 1080 |
| OY | 1213 | TAONS\$H\$E\$F\$V\$AD\$OG\$NN\$F\$V\$N\$P\$Y\$S\$Y\$D\$T\$K\$SD\$V\$T\$F\$AG\$NG\$IT\$T\$Y\$K\$N\$G\$V\$AV\$G\$D\$ | 1272 |
| Dd | 1081 | ta\$ng\$ll\$e\$fr\$vad\$eg\$nhl\$fv\$sn\$py\$sd\$tk\$sd\$vl\$ta\$gen\$g\$lt\$tk\$nk\$yv\$vv\$g\$ld\$g | 1140 |
| OY | 1273 | T\$K\$G\$LT\$P\$RL\$T\$V\$G\$N\$N\$G\$K\$G\$I\$D\$S\$O\$N\$O\$M\$T\$IG\$T\$S\$N\$T\$LA\$V\$T\$MD\$K\$S\$V\$R\$T\$E\$G\$K\$T\$R\$E\$D | 1332 |
| Dd | 1141 | tk\$gl\$tp\$kl\$lv\$gm\$ng\$iv\$ids\$kd\$gn\$tl\$gt\$sn\$tl\$an\$vt\$nd--ga\$gh\$als\$g--land\$ | 1198 |
| OY | 1333 | KTR\$AS\$IV\$D\$V\$L\$AG\$F\$N\$Q\$NG\$E\$A\$V\$D\$F\$V\$T\$Y\$D\$V\$N\$E\$AD\$G\$N\$AT\$T\$AK\$Y\$T\$D\$D\$T\$K\$T\$S\$K\$V\$Y\$V\$D | 1392 |
| Dd | 1199 | ktr\$as\$ig\$dv\$lna\$g\$ln\$g\$gn\$eav\$df\$vs\$ty\$dv\$df\$dg\$nat\$kt\$ak\$ty\$dd\$ts\$ks\$kv\$y\$dv | 1258 |
| OY | 1393 | NVD\$T\$T\$E\$V--K\$D\$K\$LG\$K\$T\$T\$T\$L\$T\$S\$T\$G\$AN--K\$F\$AL\$S\$N\$AT\$D\$D\$AL\$V\$K\$AS\$D\$Y\$AH\$N\$T\$IS\$G | 1449 |
| Dd | 1259 | nvd\$nt\$lev\$sd\$sk\$lg\$vk\$tt\$ll\$kt\$sa\$ng\$at\$F--sa\$dg\$dal\$lv\$ks\$dl\$at\$hn\$tl\$ng | 1315 |

| | | | |
|-----------|---------------------------------------|---|------|
| QY | 1450 | DIORFKGSAQNSAGVYDDADGNKKVITDSDNKKVYQAAKDGTVDTKREYAKRLVAQAKT | 1509 |
| Dd | 1336 | diqrakagssasaaasyvdaadgmkvlydsckkyyvynvdkqydknkevakklyvaqct | 1375 |
| QY | 1510 | PDGTLAQMWNVSVINKBOVDANKKQGINNDNAFVAGLEKASDNKTKNAATVGDLLAV | 1569 |
| Dd | 1376 | pdgtlaqmwvsvinkbovdankkqgineednaflglenaekdcktknaavtygdlnav | 1435 |
| QY | 1570 | AOPELTPAGbTGTTAKKLGETLTIKGGQIDTNKLTDDNNIGVAVGTDFVTKLADLNLN | 1629 |
| Dd | 1436 | apltlftagdtgtctakkrlgetltlkggqctdnltldnnlignvavgtdfvtklakoltln | 1495 |
| QY | 1630 | SVNNGCTTIDDKGVSFPDSSGQAKNPVLSANGDL | 1666 |
| Dd | 1436 | svnngcttlddkgvsfydanggakantpvlsangldlgkrlsnlgaavddndavnfkqf | 1555 |
| QY | 1667 | ----- | 1666 |
| Dd | 1556 | nevaktvnnlhngsnsgaslpfvvtadangkprlntgtqkpkaklqadgkyyhanangvpy | 1615 |
| QY | 1667 | ----- | 1666 |
| Dd | 1616 | dkdgpvltdacklanlaahgkpldaqhqvvaslgnssdaltlnlksrlpqldtptgna | 1675 |
| QY | 1667 | ----- | 1666 |
| Dd | 1676 | nagqagslpslsaagssnaasvkdvlnvgfnlqtnhngvfykaydtyvnfyngtgadlts | 1735 |
| QY | 1667 | ----- | 1666 |
| Dd | 1736 | vrsadqlmsnltvntlaaldoddggnvllakadgkfykaddlmpngslkagksaadaktpt | 1795 |
| QY | 1667 | ----- | 1666 |
| Dd | 1796 | glslvnpnagksgstgdavalmnlskavfkskdgttttvsdgsislgkdhnsiltlskdg | 1855 |
| QY | 1667 | ----- | 1666 |
| Dd | 1856 | invggkylsmngkytkctkdtaanvgqlnevrnllylqnaagndadngvnladlkkdpnsg | 1915 |
| QY | 1724 | SSSNRTVYKAGTVYGGKNDTKEKLTATGGVQGVDPDKDGNANDLSNWVWKTKQKDSKKAL | 1783 |
| Dd | 1916 | sssnrtvlykagtlvgkygmndteklavgyvqvavdgdnaangdlssnvwklykdgskkal | 1975 |
| QY | 1784 | LATYNAAGQTYVLTNNPAAEAIIDRINERGIRFEFHVNDGNQPEPVQORNGIDSSASGKHSVA | 1843 |
| Dd | 1976 | latynaagqtnyltnnpaeaidltnegqlrlfryvndngpepyvgrnglidsaagkhsva | 2035 |
| QY | 1844 | IGFOAKADGEAAVAIGHQTOGAGNOSIAIGDNMAQTDQSIAIGTNVVAAGKHSAGIDPS | 1903 |
| Dd | 2036 | lfgfakadgeaaavaighrtqdgagnsialdgnaqatbdgsialgtgnvrtgkhsagidgps | 2095 |
| QY | 1904 | TVKADNYSVGNONNOFPDANOTVDFGVGNNTTYTESNSVALSSNSAISAGTHAGIOAKKS | 1963 |
| Dd | 2096 | tvkadnysvgnnmqfidaqtavfyvgmnlvttesnsvalysnasaagthagiygaks | 2155 |
| QY | 1964 | DGTAGTTTGATGTVGAFAGATAVAGAVSAGSAGBARRIONYAAGEVATSTDAVNGSOL | 2023 |
| Dd | 2156 | dgtagtttgatgctvgvfagfagvavsvyagsgaerriqnvaagevatastdavngsqll | 2215 |
| QY | 2024 | YKATOGIANTNIELDRIHONENKANANGAISAMAMASMPDAYIPGRSVYTGCIATHNGOG | 2083 |
| Dd | 2216 | ykatcgilanaetneidhrlbgenkanaagissamamasmpdaylpgmsvtygtialthngqg | 2275 |
| QY | 2064 | AAVAGLSKLSNDNGWVWKINGSADTGGHGAAGVAGFHR | 2122 |
| Dd | 2216 | avavglsklsdngwvfkingsadtgghvgaavagfhr | 2314 |
| RESULT: 8 | | | |
| AAAR99393 | | | |
| ID | AAAR99393 standard; Protein; 2353 AA. | | |

AC AAR9393;
 XX 15-JAN-1997 (first entry)
 XX Haemophilus adhesion protein HA2.
 DE Haemophilus adhesion protein HA2.
 XX Haemophilus adhesion protein; HA2; hsf protein; vaccine.
 KM Haemophilus influenzae type b strain C54.
 XX Haemophilus influenzae type b strain C54.
 OS WO9630519-A1.
 XX 03-OCT-1996.
 XX 22-MAR-1996; 96WO-US04031.
 XX 24-MAR-1995; 95US-0409995.
 XX (UNSL-) UNIV ST LOUIS.
 PA (UNITW) UNIV WASHINGTON.
 XX Barenkamp SJ, St Geme JW;
 PI WPI; 1996-455364/45.
 DR N-PSDB; AAT41476.
 XX Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in
 PT vaccines against H. influenzae infection.
 XX Claim 5; Page 66-73; 120pp; English.
 CC Haemophilus adhesion protein HA2 (AAR9393) is associated with the
 CC formation of surface fibrils involved in adhesion to various host
 CC cells; it is also referred to hsf (Haemophilus surface fibrils).
 CC Its amino acid sequence was deduced from a genomic DNA clone
 CC (AA41476) derived from Haemophilus influenzae type b strain C65.
 CC Large quantities of recombinant HA2 can be produced in transformed
 CC prokaryotic or eukaryotic host cells, for use in vaccines against
 CC H. influenzae infection.
 CC
 XX Sequence 2353 AA:
 SQ
 Query Match 11.1%; Score 1185.5; DB 17; Length 2353;
 Best Local Similarity 22.7%; Pred. No. 1.6e-42;
 Matches 613; Conservative 330; Mismatches 832; Indels 925; Gaps 116;

DB 231 -----esvdlsaynnveftgdkntldvltakengktlevkfipkts-----Vlkek 279
 QY 402 DLKVLQAEFTLSLKWTYTNLTANKEVYVGTTRLTDRKIGFINDMNGIDESKPYLDKDF 461
 DB 280 dg-----kltfgkennldnkvlt--snaldn--ldegngltakavid-- 318
 QY 462 GHAGGOKITKLTAGVDDDAATYVQGLKKVNOJAEASALQTFYVKKVKNNGN-----DA- 514
 DB 319 avnkagvrvklttangngdfalva--sgtnvtfesgdgtlaavtkdngnltvkydak 376
 QY 515-----NDSKITIVGKNNKPDGTOYNTL-----KLGNGVDVTT----- 548
 DB 377 vgdgllkfdsdkkivadtaltvtgkvaekadkkkllvnaqdlvvalgnlswakaea 436
 QY 549 ETNGTYTFEGNONGLTIVGSLTINDGLSVK-----NTNSKQIQVAGDGTFTFDISNKP 604
 DB 437 dtdgale-giskdqvavageltvfkagknlkvkqdgantfyslqdalgtltsltlgtn 495
 QY 605 GAGIENTRTTRDGIGF--ANNTGSLDANKRRLPTGINSAGKELTNVQSAINPATWGGQ 662
 DB 496 g-gndaktvlnkgdltltspagngltgntlsrvtkdglkagkaltvvaaglr-ayddan 553
 QY 663 IDFMNRLST-----ANTE--KSSAATIKDLYNLQVPLTFAG- 698
 DB 554 fdvlnsaatdnrhvedaygjlInlneknankqplvctsaatvgydlrkigwvstkngt 613
 QY 699 DTGPNYTKKIGELLKVKVGGTTADDLTKNNIGVADSDTSLVYKLAKTISDL-----D 752
 DB 614 keesngkqgadevltfagaatvtskseng-----khlvsvaelkadgglekdgd 665
 QY 753 AVNTKTTLTA-SDKYTVDSGNNTAKLONGLTFESKONTGAPPA----- 793
 DB 666 tlrlkvngntdnvltvgmngtavtsgy--fevrvkgadadagkytvdatandadkk 722
 QY 794 -----TNSKITGVDLKFTDN--NGIALDGTITTKKVKGFKKOD 831
 DB 723 vatvkdvalnsaatfvktenlttsidednplndgkdalagkdltlfkagknlkvkrd 782
 QY 832 G-----SIDKSKPYLDK-----DKLVG-----EVELTTNGIN-----AG 861
 DB 783 gkntfdlaknlevktakvskdltlignrptgltarcpkvnltsaaglnfaketaadag 842
 QY 862 GKAI--TGLSNTLTD-ATNATGHTVQLGIVSDTKTRAASIGDVLAGFNKNGDPAKD 918
 DB 843 sknyvlkgitltltpeagkskshvdl--nvdaakksnaasleevlragmngignnvd 900
 QY 919 FVSTYDTVDYF----- 929
 DB 901 yvaltydvnftdstgtltvtvlgkadgkadvkigaksvikdngkltfgykldkdam 960
 QY 930 ----- 929
 DB 961 gatlvsedgkdtgtlvtaktvldavnksgrvrvtgegaetelatavnaaetvts 1020
 QY 930 -----NGNATTAKYVYDDKASKAVAYDVNV-DGTTIHLTGADGNKNOIGVTTTLTKTDK- 983
 DB 1021 vnfknngnatcatvskdnglnvkvdvngvggllk-----gdakklvadtltltvsgkv 1074
 QY 984 -----GDKAINEVSNAGDCKALINAKDIDNINLTAGEIR-----NTKGTADALQ--- 1029
 DB 1075 svpaganavn-----nnkklvnaeglatalnltswakadyadgesegeltdevkagd 1128
 QY 1030 -----TFQV-----KKVEN----- 1039
 DB 1129 kvtkagknlkvxqsekdfyyslqdlgtltsltlggtangrntgvtvlnkdltltlan 1188
 QY 1040 -----GDDNDADTTFYTKD-----AKTNQV 1060
 DB 1189 gaaagtlaasngntlsrvtkdglasgnkeltnvkaalktykdtgnratetqgkefnaavkna 1248
 QY 1061 NTLKLRKNGLDIOFTNKDG-----TVTFGINTQSLKAGKNNTTLNNNG-----LSIKNTAGNE 1113
 DB 1249 nevefyvgkngatvsaktcdngkhlvtldv--seakvgdglektdgkiklkydnldgn 1305

QY 2006 AAGSVAATPTDVAAGSGLYKATGCGIAN---ATNEDHDIHONENKANANGISSAMAMASMP 2062
 Db 2234 aagelssastclainsgslayavakgvtlnaagvynnlsgkvkvkrkradqatsalaasglp 2293
 QY 2063 GAYIPGRSMVYTGATATHNGGAGAAVAVGSKLSDNGQWFKINGSADTQGHVGAAGAGFHF 2122
 Db 2294 qatmpgksmvtlaagsssyggnglaigvrtisdngkvlrlrsgltinsgkltgvaagvyqvw 2353
 RESULT 9:
 AAB23860
 ID AAB23860 standard; Protein: 2411 AA.
 XX AAB23860;
 AC
 XX 17-JAN-2001 (first entry)
 DT
 XX
 DE Haemophilus influenzae adhesin (Hia) protein from type c strain APL.
 XX
 KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
 KW non-typable Haemophilus influenzae; antiinflammatory; auditory;
 KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;
 KW diagnosis; immunogenic; antigen.
 XX
 OS Haemophilus influenzae.
 XX
 PN WO20005191-A2.
 XX
 XX 21-SEP-2000.
 PD
 XX 16-MAR-2000; 2000WO-CA00289.
 PF
 XX 16-MAR-1999; 99US-0268347.
 PR
 PA (CONN-) CONNBUGHT LAB LTD.
 XX
 PI Loosmore SM, Yang Y, Klein MH;
 DR WPI: 2000-618897/59.
 DR N-PSDB; AAA92499.
 XX
 PT Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
 PT use as antigens and vaccines and for treating Hemophilus influenzae
 PT infection
 PS
 PS Claim 1; Fig 24; 275pp; English.
 CC
 CC The present sequence represents a Haemophilus influenzae adhesin
 CC (Hia) protein from the type c Haemophilus influenzae strain APL.
 CC Hia genes and proteins have antiinflammatory, auditory and antibacterial
 CC activities, and can be used in the production of a vaccine. An
 CC immunogenic composition comprising an Hia gene, a polypeptide encoded
 CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
 CC protection against disease caused by Haemophilus strains in a
 CC susceptible host, preferably a human. An Hia protein is useful as an
 CC antigen, in immunogenic preparations including vaccines, as a carrier
 CC for other immunogens, and in the generation of diagnostic reagents. Hia
 CC is useful for treating diseases caused by the infection of Haemophilus
 CC influenzae such as meningitis, epiglottitis, septicemia and otitis
 CC media. Recombinant production of Hia favours high recovery of the
 CC protein compared to the low recovery of native protein from Haemophilus
 CC influenzae species. A truncated protein has a significantly higher
 CC amount of recovery than a full-length protein.
 CC
 CC
 SO Sequence 2411 AA;
 Query Match 10.9%; Score 1169.5; DB 21; Length 2411;
 Best Local Similarity 22.6%; Pred. No. 8e-42;
 Matches 618; Conservative 331; Mismatches 855; Indels 925; Gaps 118;

Db 1 mklfnvlnwmtqlwvsvseltlthck--rasatvetavlallfatvganat----- 52
 QY 61 LINGSYAGIGISEADGCGGAGANARBGDKSIAIGDIAQ-----ALGSQSAIGDINKIVHNSN 115
 Db 53 -----dedeqlapvrtapvlsfnsdksgtgekevtensn 87
 QY 116 NNANIGAKASGNESIAIGDVLASGASIAIGSDLLKKTETVOQISELPIIRGOKALN 175
 Db 88 wglfymhg-----vfkagatlckag-dnlkikgkgt----- 117
 QY 176 DIYQLADTNLQKRYRTHAOGHASTAVGAMSTYAKGHFSNAFGTRATAEGETYSLANGLT-AT 234
 Db 118 -----nassf-----tyslkkdlldt 134
 QY 235 AKASSIAVGSNAQALGFAATVAG-----GSTOVNLNRGIALGFGSOVLQKDNVDNMAN 288
 Db 135 svateklstfgangdkvdltsdanglklaktgngvhl-----gldstclpdavtctyvis 189
 QY 289 VAVAPADDNQPIDNRKYKATFKNGADVFESIG-NSNGNDSIRKRIINVGASADTDANVA 347
 Db 190 ssstfnd-----vekratv-----dvlnagwnlkgt-----aktagnv- 226
 QY 348 QLEAVRL--ANRQITFEKGDSN-----RVEKGLCKTLITIGGAQTSALTHNIGVONG 401
 Db 227 -----esvdlvsaynveflgtcknldvvlakengktlevkftpkts-----vikex 275
 QY 402 DGLKQOLAEFTLSLKMVTENTLANKKVYGVKTRITTDKIGFTNMNGIDSKPYLKDIT 461
 Db 276 dg-----klftgkenndcnkv-----sntatdn-----tdeagnjlvakavid--- 314
 QY 462 GHAGGOKITKLTAGVDDDAATYGOLOKKVNOJASALOTFTVKKVDKNGN-----DA- 514
 Db 315 avnkgawrktltangngdfatva--sgtnvtfesgdtltasvcklengngltkygak 372
 QY 515 -----NDSKIIT-----VGNKNKPDG-----TOVNTKLKGENGV 544
 Db 373 vgdgklfidsdkklvadltaltlvrggkvaekdkkklvnagdltvltalgnlswakaea 432
 QY 545 DUTTERNGVTGFLGNONNGLTGVNSLTNNNDGLSVK-----NTNSNKOIOGADGIFFTDIS 600
 Db 433 d--tdtdgale-glskdegvkgagetvltfagknlkvkgdgnafvtslqdalqylsticlg 489
 QY 601 NSRPGAGIENTRITRDGIGF--ANNTGSLDANKPRLPTGIMAGKELTNVQASINPAT 658
 Db 490 gtlng-gndaktvinkdgtlcpagngtfgtntlsvtckdgikagnkaltvnaagldr-ay 547
 QY 659 NGGQDPMRLST-----ANTE-----KGSASATIKDLYNLQOVPJT 695
 Db 548 ddantfvlinsatdlnrhvedaykgjlnlneknankqplvtidstaaltvgdliklgnvvt 607
 QY 696 FAG-DTGPVWTKKLEIILKVKGKGTADDLTKNNIGVADVSDNSLTJVLAKTLSD--- 750
 Db 608 knglteesngvkgadevltltgagaatvtskseng-----khltsvvaetkdsagje 659
 QY 751 -----LDNAVTKTLTASDKVTVDSGNNTAKLQNGDLTFESKONTGATPA----- 793
 Db 660 kdgdtklklvndqnt-----dnvltvngngtavekkg---fetvktgtdadrgkvtyk 710
 QY 794 -----TNSKTIIGVDSLKFETDN---NGIALDGETTYT 821
 Db 711 datandackkvatlvkdvatalnsaalfvktentltsidednptdngkddalgaagdlftk 770
 QY 822 KOKVGAPODQ---SLDCKSPYLDK-----DKLKVG-----EVEITTINGIN 859
 Db 771 agknlkvrtidgknltfdlcknlekvktaksdcltltgntpbtggtatckpvnltstadaqln 830
 QY 860 -----AGGRAI--TGLSNTLTD-ATNATGHTVQLOIGVSDTKTRAASIGDVLINAGF 908
 Db 831 faketaasgsknylkgjatltiltpepagakashval--nvdatkksnaasiedvltvragw 888
 QY 909 NLKNNGDADKDFSTYDTVDFI----- 929
 Db 889 nlgngnmdyvaltydvtlftdtslgtltvtlvtkadgkvadk.lgaktsvikdngklf 948

QY 930 ----- 929
 Db 949 tgdklkanngatvseddgtgtglvtaktvldavnskgrvrtgagatavag 1008
 QY 930 -----NGNATTAKYTYDGKASKVAVDVNV-DGTHILTLGADGKNNOIGVNT 974
 Db 1009 naetvtsatvsfnkgnattatvskdnglnvkvdydvngdglkl-----gddkklvadt 1062
 QY 975 TTTLEKTDK-----GDKAIFSVNSGDDKALINAKDIADLNLTLAGEIRNTKGTADTALQ 1029
 Db 1063 tcltvvgkvsvpagansv-----nmkklvnaeglatlmlstakadk-yadgeee 1115
 QY 1030 TFQYKRVKENGDDDDNDADITTVGKDAKTNO-----VNTLKL-----KGKNG 1070
 Db 1116 getdgevka-----gdkvlttkagknlkvkqsekdftvyslqdtltglstlglgtangrnd 1170
 QY 1071 LDIOFNKDG-TVTFGINTVSGLKAAGNTTIN--NNGLS-----IKMT 1109
 Db 1171 tglvinkdgtliltlangaaagtdasngntlsvektkdsagnkeltnvksalktykdqnt 1230
 QY 1110 AGNEQIQVGA-----DGVKFAKVNNGVAGIDG-----TTRIT-----LSKDG-- 1143
 Db 1231 agatvpaanraevakqdlvdltkpatgaagngadakapdtlaatvlgdrlgvlwlsakt 1290
 QY 1144 -----RDEIGFAGTNGS-----LDRSKPH----- 1167
 Db 1291 adetqkfehaavknaneveftvgngatvsaktcdnogkhlvtldvaeakvgdglektdg 1350
 QY 1168 -----INAGGKKTINIGSELTAQNSNDNAVTC-----GRTY-----DL 1199
 Db 1351 kklkvlndtgnlltvdaktgaasvafnvtldteltgagthuanergvkvksngat 1410
 QY 1200 KTELENKISSAKTAQNSLHEFS--VADEQGNFTVSNPVS---SYDTSKTSVITF-AG 1253
 Db 1411 atetckkkvatvgdvakaindaatfkyvenddsatltdspddgandalagdtliltkg 1470
 QY 1254 ENGITTIVKNGV---VRVGIDQTKGLTTPKLVGNNGK-----GIVDSONGON 1300
 Db 1471 knlvkrdgnkltfalandlsvksatvsklsjltngknltscdkglnfakdsrtygd 1530
 QY 1301 T-----ITGLSNTLAVNTDKGSVRTBOGKILIKDDKTRASIVDVLSAGPNLOG----- 1351
 Db 1531 anhlnglaestltdltlinsat-enlsgngltdeektraasvkvlnagvngvkvpsa 1589
 QY 1352 --NGEAVDFVSTYDTPVNEFADGNATTAKVTYD--DTSKTSKVVVDVNVDDTYIEVKDKKLG 1407
 Db 1550 ngyenldfvtatvtdtvsdgkdtstvtveskngkrtev-----kig 1633
 QY 1408 VKTTTLTSTG---TGAN-KFALSNOAT-----GDALVKASDIIVAHNLNLSGDIOT 1453
 Db 1624 aktsvikdngklftgkelkldannngvltetdgdkgenglvtaekavldavnkagvrvlt 1693
 QY 1454 AKGASQAN-----SSAGYVDADGNKVITYDSTDNKYYQAKNOS--HYDATKEVAK----- 1500
 Db 1694 t-gangngndatavagstuvlfadngltlaevt-----kandslvkvnvkvadglkl 1746
 QY 1501 --DKLVAQA---QTPDGLAOMNVKSVINKEOVDANKKOGINEDNAFVGLKEKASDNK 1555
 Db 1747 dgdklvadtvtlvvadgkvrtapn-----ngdck-----fvaaagjlaadlnk 1788
 QY 1556 TKMAAVTVGDINAVAQPLRTPAGDTGTTAKKLGFTLTIGSGQDTFNKLTNNIGVAGTD 1615
 Db 1789 lswta-tagkegtgevpdpsansag-----evkagkvltfkag-----dnlkikgsgk 1834
 QY 1616 GFTYKLANDLNLNSV-----NAG-----GTRIDDKGVSPFSDSGQA-----KANTPYLSA 1661
 Db 1835 dftyslkelkeldtsvefkdaangltgsestkltdgtltpangagaaganatavtsvck 1694
 QY 1662 NGILDGKGVISNKGKGT-----DTDAANYQOL--NEVRMLLGLGNAGNDN---A 1706
 Db 1895 dglsagknavtnvsglklfkgdghtlangvtdvdekhynadnyakeltnldekdagnpntva 1954

Example6: Fig.6: 646pp; English.

Db 585 aagaafsaahgtsahtskl-----tunvag-----nltagsida 616
 Qy 658 TNGGLODFPMNLSTAPNNEKSSSATIKDLYNLGVLFPAGDPCPNVTKRLGELAKKGG 717
 Db 617 vngsqllktndnvtcnc-tlnactnltlnltda-vnlgjdsdlwnktagafsaahgt 673
 Qy 718 KTTA--DDLTNNNGVAVADSTDNLSLFWLAKTLISDLDAVWTKPL-TASDKVWVDSGN--- 771
 Db 674 datsklnvtaagnl-----tagstдавngsqkktndvntctnhtiat 716
 Qy 772 NTAKLQNGDLTFESKQNTGAPPAINSKITIGVDGLKFTDNGNIALDGTYYITKDKVGEAKOD 831
 Db 717 nttitn-ltdavngjdsdlwnktaga-fsaahg--tdatsklnvtagadlta- 767
 Qy 832 GSLDKSRPYLDKOKLKAGEVEITNGINAGKATGTSNLTATNA-----TTG 881
 Db 768 gstda---vngsqllktndnvtcnc-----tlnlt-n-ltdavngjdsdlwnktag 814
 Qy 882 HVTOLGIVSDTKTAAISIGDVLNAGFNLKNNGDAPFVSTYDVPDFINGNATTAATYTD 941
 Db 815 afsaahgtatsklnvtagad-ltag-----stдавng--sqllktnd 854
 Qy 942 GKASKAVAYDVNWDGTTIHLTGADGNKNQIGVKTTLTK-TDANGDKAINEVSNSGDKAL 1000
 Db 855 -----nvtntnltlnltdsvgdtk-----dssll 879
 Qy 1001 INAKDIADNLTLAGETIRNTKGT-ADTALQTFQYKKVKEGDDNDMDITTVGKDKATNQ 1059
 Db 880 -----wnkaagaafsaahgtsahtsklnltagkissnstdalngsqlygvadetsy 930
 Qy 1060 VNTLKLKGNGLDIQTKNDGTVTFGINTQSLKAGNNTTNNGLSIKTNAGNEOIOVGA 1119
 Db 931 lg-----sgadt-----sdtgvsjsgplytg----- 951
 Qy 1120 DGVAFAYNNNGVYAGIDGTTTRITRDEIGFAGTNGSLDKSKPHSKDGINAGKKTITNIQ 1179
 Db 952 -gtclvngvalaaaintsfstl-gdalnwdatagkts-----akhnginapsviltva 1003
 Qy 1180 SGEIAONSNAVNTGKTYDKTLEKNKISSFTAKTAON---SLHEFSYADQGNNTFYNSP 1236
 Db 1004 ngavststsdalngsqlygvadyadalggnavntgtsiltcpltyaaggsyn--vgda 1061
 Qy 1237 YSSVYTSKTSQDVI---TFPENG-----ITTKVNGVAVF---GIDOTKGL 1276
 Db 1062 leadtlddallahttangngafsaahgkdktaasvltvanavatsndalngsqly 1121
 Qy 1277 TTPKLTLYGNNNGKSTIVDSONGONTITGLSNTLANVT-NDKG-SVRTTEOGKTIKDEK 1334
 Db 1122 stnkyiadaalvgdae--nadgltaptylantdyinnvgaaldaalnalwged-- 1175
 Qy 1335 RAASIVDVLASGFNLGNGEAVDF-----VSTYDTVNFADGNNTTAKVYYDDTSKT 1385
 Db 1176 -----agaynaashdnaskltnvaagdlstscdavngsqlnatnltv-----qn 1221
 Qy 1386 SKVVYDV--NVDDTTIE-----VKDKLGYKTTTLTSTGANKF---ALSNOATGD 1432
 Db 1222 sqmnlngagtseyleengaglnyvtlndsglafindasssglatavynavaassav 1281
 Qy 1433 ALKVASIVAHNLNTLSGDIOTPAKASQANSAGYVADAGKVIYDSTDNKYAKKDGVY 1492
 Db 1282 al--gqdsisevdt---glaigsssvss-----rvlwkgltn----- 1313
 Qy 1493 DKTEVEVAKDKLVAAQOTPDGTLAQMNVKSYINKEOVDANKKOGINDNMFVGLKKA 1552
 Db 1314 -----tsveegvnygdltldgel--lgaist-----gddgkyrqln-----va 1351
 Qy 1553 DNKTKNAVVVGD---NAVAOTPLTF--ACDPTGTAKKLETLTKGOTDTNKLTDN 1606
 Db 1352 dgsaahavavtrvqmqnagivatpcklymanstaedsavgedslamgakt-----lvg 1407
 Qy 1607 NIGVAVGTGCTYVAKLADLTNL-----NSVNAGSTKIDDKG--VSFVDSGQAK 1653
 Db 1408 nagjgltn--tlvlatdnglaigsnaranahadslamngsgttrgaqcnlylaymdap 1465

Qy 1654 ANPVLASGLDLCGKVISNVGKTRDPDAANVQOLEVRNLGLGNAGNDMAQNOYNI 1713
 Db 1466 qnsvgefsvgeaqqqlunvaagsadtdaenvvqg-----kvtld 1505
 Qy 1714 ADIKDPNSGSSNRTYVAGTVLGKGNNDTEKLTATGVGVVDKDGNDANGLDSNVWK 1773
 Db 1506 aqvsqngtstlnltngvt-----nldtr-----ven:engldlv----- 1540
 Qy 1774 TQKDSKKALLATYNAAGTNTLTNNPAEADIRINEOGIRPFHVNDGQOEPPVVGCRNID 1833
 Db 1541 -----tsgstkyfktn-----tdgad 1556
 Qy 1834 SSASGKHSVAIFQKADDEAVALIGROTQAGNOSIAIGDNQATGDOSIAGTGNVAG 1893
 Db 1537 aaqgkdsval-----gsgslaaadn-----svaigtvsv--- 1586
 Qy 1894 KHSGLGDPSTYKADNSYVGNNOFTDQTVDFGVGNITVYESNSVA-LGSNSAISA 1952
 Db 1587 -----adeentlsvgs-----tngrltcnvaagnatdaenvvqkxseaa--g 1628
 Qy 1953 GTHAGTQAKKSPGCTAGTTTATGATGTWGFAGOTAVGAVSVGASGAERRIONVAAGEVSA 2012
 Db 1629 gyrydtkadgsldysnltlvgngsglt-----rtsnvsag---v 1664
 Qy 2013 TSTDAVNGSOLYKANGGINATN---ELDHRIRHONENKANAGISSAMAMASMPQATYIG 2068
 Db 1665 nmddavnyqllqsvqetkytdqrmvendnklsktsklsagslasamamqllpqaylprg 1724
 Qy 2069 RSWTGTGATIHNGOGAVAVGLSKLSDNGQWFKINGSADTQGHVGAAGAFHF 2122
 Db 1725 asmasjggtlygesavalgsvmsangrvwykllgstnsqgeysaalqagjqw 1778

RESULT 11

AAB23856.
 ID AAR33856 standard; Protein; 1104 AA.

AC AAB23856;

DT 17-JAN-2001 (first entry)

DE Haemophilus influenzae adhesin (Hia) protein from NTH1 strain 29.

KW Hia: adhesin; Haemophilus influenzae adhesin; NTH1; infection; vaccine;
 non-typable; Haemophilus influenzae; antinflammatory; auditory;
 KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;
 diagnosis; immunogenic; antigen.

OS Haemophilus influenzae.

PN WC200055191-A2.

PD 21-SEP-2000.

PF 16-MAR-2000; 2000WO-CA00289.

PR 16-MAR-1999; 99US-0268347.

PA (CONN:) CONNAUGHT LAB LTD.

PI Locsmore SM, Yang Y, Klein MH;

DR WPT: 2000-618697/59.

DR N-PSDB; AAA92495.

PT Novel nucleic acid encoding Haemophilus influenzae adhesin protein, for
 use as antigens and vaccines and for treating Haemophilus influenzae
 infection

Claim 1; Fig 20; 275pp; English.

CC The present sequence represents a Haemophilus influenzae adhesin (Hia)

protein from the non-typeable Haemophilus influenzae (NTHi) strain 29.
 CC Hia genes and proteins have antiinflammatory, auditory and antibacterial
 CC activities, and can be used in the production of a vaccine. An
 CC immunogenic composition comprising an Hia gene, a polypeptide encoded
 CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
 CC protection against disease caused by Haemophilus strains in a
 CC susceptible host, preferably a human. An Hia protein is useful as an
 CC antigen, in immunogenic preparations including vaccines, as a carrier
 CC for other immunogens, and in the generation of diagnostic reagents. Hia
 CC is useful for treating diseases caused by the infection of Haemophilus
 CC influenzae such as meningitis, epiglottitis, septicaemia and otitis
 CC media. Recombinant production of Hia favours high recovery of the
 CC protein compared to the low recovery of native protein from Haemophilus
 CC influenzae species. A truncated protein has a significantly higher
 CC amount of recovery than a full-length protein.

XX Sequence 1104 AA;

Query Match 9.5%; Score 1018.5; DB 21; Length 1104;
 Best Local Similarity 27.8%; Pred. No. 9.1e-36;
 Matches 368; Conservative 165; Mismatches 399; Indels 393; Gaps 53;

QY 943 KASKVAVYDVNDGTHLTCAGDGNKQIGVTTTLTKTAKDKAIFNSVSGDD----- 997
 DB 28 kasatvaavaalataalsat-aeann-----tsvtnghlaygdtnfntnmsiadekhv 81
 QY 998 ----KALINAKD-----IADNINLTAGEIR-----NFKGTADTALOTFOYKKAYK 1037
 DB 82 qdaykglhlnnekdnksfllvadnlaetvgnlrklygwlskngyrne--ksyqkqgd 139
 QY 1038 ENDDDDNDADTIVGKDATNQNTLKLKGNKGLDIQTKKDGIVPFMGITQSGLKAGNNT 1097
 DB 140 e-----vltsggaatvs-----ssskdghhtltstvkysfaevxt 176
 QY 1098 TLNNGSLIKNTAGNEQIOVGADGVFAKVNNGVAGIDGTTTRTRDEIGFAGTNGSLD 1157
 DB 177 dactgg-----gvnadrgkvaeden--gadvd----- 202
 QY 1158 KSKPHLSKDGICINAGKKTINIGSGEIAONSNDNAVGTGKIYDLKTELENKISSAKTAQNS 1217
 DB 203 -----kkvatvk--dvakaindaetfvkystdddiengaaqghnettdga 245
 QY 1218 LHEFSVADEGNNFTV---SNPYSSYDTSKTSVITFAGENGITTVKNNGVAVGIDQTK 1274
 DB 246 l-----kagdtcltlkagknlkakiqngkvytfaialakoldvtasakvskisigkdnk 298
 QY 1275 -----GLTTPKLTIVNNGNGKGIYIDSONGONTITGTSNTLIANTNDKGSVRTTEQCK 1326
 DB 299 vdltsdangklak--tgnng-----qngvnhlnglastl-----tdltlgtmtqaan 345
 QY 1327 TIIDEDKTRAASIVDYLASGFNLOGEAVPVSYDYVAFADGNAITAKVITYDDTSKTS 1386
 DB 346 gvaevghnraasavadvlnagwlnqngasvdfvnaaytdvfvngtlnvnttdtshkt 405
 QY 1387 KVVYVD-----VAVDDTTEIVK-----DKTKG-----VKTTLTSTGSG 1419
 DB 406 tvrvdvtglrvqvtedgktvrvdnkyyeakqdsadmkkvengelaaktkvksasg 465
 QY 1420 ANFEALSNQATGALYKASDIVA--HLNLTSGDIQTAKGASQANSAGYVADGNKVI-- 1475
 DB 466 qnpvklstnvaeg---teendavsfkqkalg-ekyvtlqasnaayanngn--dagqkatct 520
 QY 1476 -----YSTDNKYYQAK--ND-----GTVDKTKVAKDKLVAAQAPDGTGLAQM 1517
 DB 521 lnnqlnfkfstgdgellnkvendtvlfpkksv--gvgedgkatlqngklttdglvaes 579
 QY 1518 NVKSVIKKEQVGNKOGINENAFYKLEKASDNKTKNAAVTYGDINAQVAGTQLTFA 1577
 DB 580 el-----veslnklgwkvvygdkgds-----gldggaenetlvks 613
 QY 1578 GDTGTFAKKIGETLTIGGOTDNNKLTLDNNGVAVAGTGTGTVKLANDLNINSV----- 1631

DB 614 gdkvvl--kagenlkvk--qdgtn-----falyakdelgvkvsvefdta 654
 QY 1632 ---NAGGKTKIDKGVSPFDSSG-----QAKNTPVLANSGLDGLGKTVISNVGKTKPTDA 1683
 DB 655 ngsngascklkcdgltlsangangaatddcklkvasdgsasgnkavknvsglkkfcd 714
 QY 1664 ANVOOLNEVRML-----GLGMNGNDADGNQVNIADIKRDPNSSSSNFTVITKAG 1734
 DB 715 anflptsadnlckkygdaykgltnldekgadqqltvd-----nt-----aa 759
 QY 1735 TVLGKGNNDIEKLATGSGVGVVDKDNANGDLSNV--WTKQKGSKALLATYNAAGQT 1793
 DB 760 tv-----gdrlglgw--lsadkltgelnkayna---- 786
 QY 1794 NYLTNNPAEALDRINEGIRFEHVNDGQEPVVOGRNGIDSS---ASGKHSVAIGFOKA 1850
 DB 787 -----gvnanevfkksq-nglhwsgktvngreit--felak 821
 QY 1851 DGEAAVAIGROTOAGNQSIAIGDNAATGDQSIAGTGNVAVGKHSAGIDDPSTV--KAD 1908
 DB 822 dena-----laigyksalrtdtvalgtgnvnaeksgafgdpnyledkag 867
 QY 1909 NSYSVGNNOPTDATQTDVFGVGN-----ITV-----T 1937
 DB 868 gsyatgdnrit--sktflvlgvnyakkykangdvdtetvtdkdkgkettvtpkalgat 926
 QY 1938 ESNVALGSNSAISAGTACGTAQAKKSDGTACTTGTAGTGKVPAGOTAVGASVGSAG 1997
 DB 927 vensvylgnk--statckkgknlksdgaqnttaagltgvtvngtaagatavsvgsag 983
 QY 1998 AERRIQNVAAGEVSAATSDAVNGSOLYKATOGIANATMELDHRIHONENKANAGISSMA 2057
 DB 984 eerlignvaageisatsdcalngsqliyavakgvt-----lagynkvykxadagtaala 1039
 QY 2058 MASPMOAVYIPGRSMVTGGIATHNGGAAVAGLSKISDNGQVWFKINGSADTQGHVGAAG 2117
 DB 1040 aeqlpqaempgkswvslagsysyqngqiaigvstrlsdngkvlrlsgltngsgktygaag 1099
 QY 2118 AGFHF 2122
 DB 1100 vgyqw 1104

RESULT 12
 AAB23859
 ID AAB23859 standard; Protein: 1104 AA.
 AC AAB23859;
 XX
 DT 17-JAN-2001 (first entry)
 XX
 DE Haemophilus influenzae adhesin (Hia) protein from NTHi strain K22.
 XX
 KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
 KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;
 KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
 KW diagnosis; immunogenic; antigen.
 KW
 OS Haemophilus influenzae.
 OS
 PN WO200055191-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 16-MAR-2000; 2000WO-CA00289.
 XX
 PR 16-MAR-1999; 99US-0268347.
 XX
 PA (CONN-) CONNACHT LAB LTD.
 XX
 PI Logsmore SM, Yang Y, Klein MH;
 XX
 DR WPI: 2000-618897/59.

DR N-PSDB: AAA92498.

XX Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
 PT use as antigens and vaccines and for treating Hemophilus influenzae
 PT infection

PS Claim 1; Fig 23; 275pp; English.

XX The present sequence represents a Haemophilus influenzae adhesin (Hia)
 CC protein from the non-typeable Haemophilus influenzae (NTH) strain K22.
 CC Hia genes and proteins have antiinflammatory, auditory and antibacterial
 CC activities, and can be used in the production of a vaccine. An
 CC immunogenic composition comprising an Hia gene, a polypeptide encoded
 CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
 CC protection against disease caused by Haemophilus strains in a
 CC susceptible host, preferably a human. An Hia protein is useful as an
 CC antigen, in immunogenic preparations including vaccines, as a carrier
 CC for other immunogens, and in the generation of diagnostic reagents. Hia
 CC is useful for treating diseases caused by the infection of Haemophilus
 CC influenzae such as meningitis, epiglottitis, septicemia and otitis
 CC media. Recombinant production of Hia favours high recovery of the
 CC protein compared to the low recovery of native protein from Haemophilus
 CC influenzae species. A truncated protein has a significantly higher
 CC amount of recovery than a full-length protein.

XX Sequence 1104 AA;

Query Match 9.5%; Score 1018.5; DB 21; Length 1104;
 Best Local Similarity 27.8%; Pred. No. 9.1e-36;
 Matches 366; Conservative 165; Mismatches 399; Indels 393; Gaps 53;

QY 943 KASKAVADVNDGTTIHLTGADGNKNOIGVTTTLTKDADKAINFSVNSGDD----- 997
 DB 28 Kcasatavavlatatlatat-aeann-----tsvtnglmgysdntftmstadlekv 81
 QY 998 -----KALINAKD-----IADNLTLAGLR-----NKGADTALQTFYQKVK 1037
 DB 82 qdaykgljlnlekdnksflvadtaevnrlkglwslskngtrne--ksyqvqad 139
 QY 1038 ENGDDDDADDTIYGNKAKTNOVNTLKLKGNGLDIGNDGYTFPFINOSGLKAGNNT 1097
 DB 140 e-----vltfsgsaats-----ssskdghkhtllsvtkgsfaevkt 176
 QY 1098 TLNNGSLIKNTAGNEIOVAGDGVKFAKYNNGVAGIGITRTTRIDEIGFAGTNGSLD 1157
 DB 177 datgg-----qgnadgkykaeden--gadvd----- 202
 QY 1158 KSKPHLSKDGINAGKKTITNIOGELIAONSNDNAVTKGRIYDLKTELENKISSAKTAQNS 1217
 DB 203 -----kkvatvk--dvakalndaatfkwkestdddlengagaqnetldga 245
 QY 1218 LHESVADDEGNNTFY---SNPYSSYDTSKTSVITTFAGENGITTKNKGVAVGVIDQTK 1274
 DB 246 l-----kagddcltlkaqknlkxldqngksvtfalakldvtsakvsklslgkdtk 298
 QY 1275 -----GTTPKLTGVNNGKGIIVDSONGNTITGLSNTLANVTNDRGSVRTTEQOK 1326
 DB 299 vdltsdanglkak--tgnng-----qngvnhngiasltl-----tdltgmmtqasn 345
 QY 1327 IINDEKTRASIVDYISAGENLOGEAVPEVSTYDVFADGNATTAKTAVYDTSKTS 1386
 DB 346 gvaqgmhnaasavdlnagwmlqngasvdfvnaydltvfnngtnlvnttdtchakt 405
 QY 1387 KVVVD-----VVNDDTTFYVK-----DKKLG-----VKTTLTSTGTCG 1419
 DB 406 tvrvdvglpvqyvtedgkvtvkvdkyyeakqdsadmkkvengelaktkkyvasag 465
 QY 1420 ANKFALSNQATGDAIVASDIVA--HLNLTSGDIQTAKGQANSAGYADADGNKYI-- 1475
 DB 466 qnpvklsnvaeg---teendavsfkqlkalq--ekyvltsasnayangn--daagykacqt 520
 QY 1476 -----YSTDNKYIOAK--ND-----CTVDKTKEVAKDKLVAAQNPDPGLAOK 1517

DB 521 lngnlfkfstldgeellnkqvendvtfcpkksv--qvgedgkatlqngtktdglveas 579
 QY 1518 NVKSVLNKEQVNDANKKOGINEDNAFVKGLERKASDNKTKMAAYVGDILANAAQPLPFA 1577
 DB 580 el-----veslnklgwkkyvdkds-----geldgasnetlvs 613
 QY 1578 GDTGTTAKKLGERTLTIRKGGQDTDNKLTLDNNIGVACTDGFYVAKDLNLSV----- 1631
 DB 614 gdvkvl--kagelnkvk--qdtgn-----ftyalldelgvxsvelfkda 654
 QY 1632 ---NAGCTKIDKGVSEFVDSG-----QAKANTPYLSAGDLGKVSINWCKGTKDDA 1683
 DB 655 nsgnsgstkltqyltltasangangaadcklkvasdgiagmkavknvvsjklkkfyd 714
 QY 1684 ANYQULNEVRNLL-----GLGNAGDNADGNVNIADIKKPDNCGSSSSNRVYKAG 1734
 DB 715 anfnltsadnltkyddaykqltnldekqadkqqltvad-----nt-----aa 759
 QY 1735 TVVGGKGNNDTEKLAGVGVGVVDKDGNNAGDLSNV--WVKTKDKDSKKALLATYNAAGOT 1793
 DB 760 tv-----gdltglgwv--isadkltgelnlkeyna----- 786
 QY 1794 NYLTNNPAPALDRINRQIRFEHVNDGNOEPVVOGRNGIDSS---ASGKHSVAIGFOAKA 1850
 DB 787 -----qvrnanevkkfksg--nglhvsgktvngtrfelt--felak 821
 QY 1851 DGRANAATGRQVAGNOSIAIGDMAOTGDSIAIGTGNVAVAKISAGIDPSTV--KAD 1908
 DB 822 dena-----lafygsakalrndeivaigtvnnvnaeksgafgdpnyledkag 867
 QY 1909 NSYVGNNOFTDATTQDTDFGVGN-----ITV-----T 1937
 DB 868 gsyafndndrit--skrtflngvnyakkyangdvdtelvtvdkdggkeltvtpkajal 926
 QY 1938 ESNVVALGSNSAISACTHAGTQAKKSDGTAGTTTATAGATGVKGFAGOTAVAGSVASG 1997
 DB 927 vensvlygnk-----statkdkylnkdsfgtagnnttgaigtvgngfagstlabavsgasg 983
 QY 1998 AERRIONVAAGEVSATSTAVANGSOLYKRTGCIANATNELDRIRIONENKANAGISSMA 2057
 DB 984 eertrignvaagelsatstidaingqilyavakgvtr-----laqgvnkvyktradaftasala 1039
 QY 2058 MASMPQAVIPGRSMVTGATATNCGOAVAVGSKLSDNGOWEFGKNGSADNPGHGAIVG 2117
 DB 1040 asqldpqaampgmsvsiagsyqgqnglaivysrlsdingkvlrlisgtmsgkvtgvaag 1099
 QY 2118 AGFHF 2122
 DB 1100 vgyqw 1104

RESULT 13
 AAB23857
 ID AAB23857 standard; Protein; 1004 AA.
 XX
 AC AAB23857;
 XX
 DT 17-JAN-2001 (first entry)
 XX
 DE Haemophilus influenzae adhesin (Hia) protein from NTH strain M407.
 XX
 KW Hia; adhesin; Haemophilus influenzae adhesin; NTH; infection; vaccine;
 KW non-typeable; Haemophilus influenzae; antiinflammatory; auditory;
 KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;
 KW diagnosis; immunogenic; antigen.
 XX
 OS Haemophilus influenzae.
 XX
 PN WO200055191-A2.
 XX
 PD 21-SEP-2000.

PF 16-MAR-2000; 2000MO-CA00289.
 XX
 PR 16-MAR-1999; 99US-0268347.
 XX
 PA (CONN-) CONNACHT LAB LTD.
 XX
 PI Loosmore SM, Yang Y, Klein MH;
 XX
 DR WPI: 2000-618897/59.
 DR N-PSDB: AAA92496.
 XX
 PT Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
 PT use as antigens and vaccines and for treating Hemophilus influenzae
 PT infection
 XX
 PS Claim 1; Fig 21; 275pp; English.
 XX
 CC The present sequence represents a Haemophilus influenzae adhesin (Hia)
 CC protein from the non-typable Haemophilus influenzae (NTHI) strain M407.
 CC Hia genes and proteins have anti-inflammatory, auditory and antibacterial
 CC activities, and can be used in the production of a vaccine. An
 CC immunogenic composition comprising an Hia gene, a polypeptide encoded
 CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
 CC protection against disease caused by Haemophilus strains in a
 CC susceptible host, preferably a human. An Hia protein is useful as an
 CC antigen, in immunogenic preparations including vaccines, as a carrier
 CC for other immunogens, and in the generation of diagnostic reagents. Hia
 CC is useful for treating diseases caused by the infection of Haemophilus
 CC influenzae such as meningitis, epiglottitis, septicemia and otitis
 CC media. Recombinant production of Hia favours high recovery of the
 CC protein compared to the low recovery of native protein from Haemophilus
 CC influenzae species. A truncated protein has a significantly higher
 CC amount of recovery than a full-length protein.
 CC
 XX
 SQ Sequence 1004 AA:
 Query Match 9.1%; Score 974; DB 21; Length 1004;
 Best Local Similarity 28.2%; Pred. No. 6.5e-34;
 Matches 337; Conservative 146; Mismatches 330; Indels 382; Gaps 45;
 QY 1073 IOTNKDGTVEGINTGSLKAGNNTLNNGLSIKNTAGNEQIOVGADGVKFAVNGGV 1132
 DB 43 lttvgtattgttstnglkaygst--nnpf--naag----- 77
 QY 1133 GAGIDGTRTRDEIGRAGR-NGSLDKSKPHLSKDGINAGGKITTINQSGEIAONSNDAY 1191
 DB 78 -----satdlarq--fdgaydjlInlnkdankn-----llvtddkaa 113
 QY 1192 TGGKTYDKTELEKISSTAKTAQNSLHEFSVADEQGNFVSNPYSSYDTSKSD--V 1248
 DB 114 tvgntlrklgwlskskngtrnksqvkhnadevlfeqkdgyt-----tsksengkht 165
 QY 1249 ITFAGENGITTKVKKGVVGVGIDQKLTPEKLTGVNN-----NGKGIYIDSQNG 1298
 DB 166 vtfleklank-----natvsklslgngqhvdlststnglkfkfksptng 212
 QY 1299 QN---TTGSLNTLANVNTNDKGSVTRTEQCKIIDEKTRASIVDVLSAGENLQNGEA 1355
 DB 213 qnguvhnglasieltldtlt--gtksatng--vdvgnlnhrasavadvlnagmngqngas 268
 QY 1356 VDFSTVDTVFADGNNATTAKVTYDDNDSKTSKVYVDVNDVDTTEVADKIKGVKTTTLTS 1415
 DB 269 vdfvntvdtvfnlgnlnvnt--tdahnkkt--vrvdvtgipvq----- 312
 QY 1416 TGTGANKFALSNQATGADLVKASDIVAHLNLTSGDIOTAKGASQANSAGVADGKNVI 1475
 DB 313 -----ytdedgetv 322
 QY 1476 YDSIDNKTYQAKNDGTVDKTEVAKDKL-----VAQACTPDGTLAQNANVK 1520
 DB 323 --kvneyeakqdsadmckkvengklakckvklsangnprklsinvadt---entd 377

QY 1521 SYNKEQVNDANKKQGINEDNAFV-----KGLE-----KNA 1551
 DB 376 avsfrklkalqdkvrlslasnayangsdsdggkqgqtlstnglnfkfstgdellnlkka 437
 QY 1552 SDNKT---KNAAYTVGD-----LNNAVQTPLPFADGTGT 1583
 DB 438 pdtvftprkksvqvvgddgkatlqdgakettglvaseelwdslnkngkvvgtdgttyt 497
 QY 1584 AKKLGELTITIKGGQTPNTKLTNNIGVAGTGFYTKLAKDITNLNSV-----NAG 1634
 DB 498 dgthcttl-vksqdkvllkagdnlkvkggeln-ftyalldeltvksvefrkdtlangana 555
 QY 1635 GTRIDKGVSPVSSGQA-----KANTPVLSANGDLGKGVSNVNGKGTDPDAANVOOL 1689
 DB 556 gkrtktdgltitlpangagaagantantltsvtkdgsagnkavknvsgjkkfgedanfdpl 615
 QY 1690 NEVRNLLGLGNAGN-----DNADGNQVNTADIKDQNSGSSNRFVYKAGTVLGKGNND 1744
 DB 616 t-----ssadnlkqydnaykgtlnldekskqkqtrptvadnt---aatv----- 656
 QY 1745 TEKLATGCVGVGDKGNANGLDSNV-WYKTQKDSKKALLATYNAAGGTNLTNNPAA 1803
 DB 657 -----gdlrglgwv-isaadklkgeInkayna----- 681
 QY 1804 IDRINQGIREFHVNDQNGEPVVGNGRNGIDSASGK---HSAVIGFOAKADGEAAVAIG 1859
 DB 682 -----gvnanewkfksq-ngl-nvsgkltlndgtreltfelekdena----- 721
 QY 1860 ROTQAGNOSIAIGDNDAGDOSIAIGTGNVYAGKHSAGIDPSTV--KADNSYSVGNNN 1917
 DB 722 -----lafsgskalrntvtaigtgnvnaeksgfsgpnyledkagsysfndn 772
 QY 1918 QFTDAGTQTVFVGNNI-----TYTESNSV-----ALG---SNSATS 1951
 DB 773 ri--tskntflvgnsvnaakrdangnvlteekveygkdgaktktvtpgalgetvensyvl 829
 QY 1952 AGTHAGTOAK---KSDGAGTGTGTAGATGVYGFAGQAVGAVSGAGARRIOTNVAA 2007
 DB 830 gnaastakdkgknlksdgcagntttagatgtyvngfagalahavsvsgaeeirngvna 889
 QY 2008 GEVSATSTDAVNGSOLYKATQIANNATNELDRIRHONENKANAGISSAMAMSPQATVP 2067
 DB 890 gelsatstdainsqlyavakgytn-----laqvlnkvkradagtasalaasqplpasmp 945
 QY 2068 GRMVTGIGATIHNGQAVANGLSLDNGQWPKINGASDITQGHVGAAGAAGFHF 2122
 DB 946 gksmvsfagsysygsqglayvrisldngkvlirfsgttnsqktygaagvyqw 1000
 RESULT 14
 AAB23854
 ID AAB23854 standard; Protein; 1002 AA.
 XX
 AC AAB23854;
 XX
 DT 17-JAN-2001 (first entry)
 XX
 DE Haemophilus influenzae adhesin (Hia) protein from NTHI strain 33.
 XX
 DE Hia; adhesin; Haemophilus influenzae adhesin; NTHI; infection; vaccine;
 KW non-typable Haemophilus influenzae; anti-inflammatory; auditory;
 KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;
 KW diagnosis; immunogenic; antigen.
 OS Haemophilus influenzae.
 XX
 PN MO20005191-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 16-MAR-2000; 2000MO-CA00289.
 XX
 PR 16-MAR-1999; 99US-0268347.

PR 19-SEP-1996; 96JP-0271408.
 XX (KAGA) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.
 PA (KAGA) CEMO-SERO-THERAPEUTIC RES INST.
 XX Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;
 XX MPI, 1998-230318/20.
 DR N-PSDB; AAV22837.
 XX
 PT Antigenic polypeptide from Haemophilus paragallinarum induces HI
 PT antibody production - and is useful for diagnosis of and preparation
 PT of vaccines for chicken infectious coryza
 XX
 PS Claim 5; Page 71-87; 108pp; Japanese.
 CC The present sequence represents an antigenic protein derived from
 CC Haemophilus paragallinarum strain C-53-47. The antigenic protein
 CC stimulates the production of HI antibodies in fowl. The protein
 CC and DNA coding for it can be used in the preparation of vaccines
 CC for the prevention of chicken infectious coryza (CIC). The protein
 CC and its antibodies can be used in the diagnosis and treatment of CIC.
 CC
 SQ Sequence 2039 AA;

Query Match 7.6%; Score 809.5; DB 19; Length 2039;
 Best Local Similarity 21.6%; Pred. NO. 1.6e-26;
 Matches 515; Conservative 301; Mismatches 830; Indels 735; Gaps 102;

QY 12 ATGTMAVAEYKSHSTGGSCATGCVSVRLSPARIALAVLYIGATLNGSAVAGIGI 71
 DB 124 aggs-iaiygagkndmwastraiaygagknes-----lasdsvaismnifkfsiviyl 177
 QY 72 ---SEADSGKGANARQDKSIAIGDIAQALQSOSIAIDNKIVHNSNNANIGAKASGNE 128
 DB 178 netyldprapbesrg--svligenaksagsgvslgpn-----awektn 221
 QY 129 STAIGDVLASGASIAIGSD-----DLYLKK 155
 DB 222 sisigagtaegkstiaigtckilgtyncklpapswdrtgkaplnsiwdfiselymgk 281
 QY 156 ETIVQOISELPIIRGOKALNDIYQOLADNLQKY-----RRTHAQG 195
 DB 282 kngcdydakkkndrpnkpeafytsdfk-sryvmpstsplyaaiklaialagstliaag 340
 QY 196 HASTAVGASVYAKGHFNSAFTRA-----TAEG---TY---SLAVGTTATAKAAS 239
 DB 341 emstavgslafadkclstamjlrstfvakdaygtalgeestrfakdsvalgnktaasnag 400
 QY 240 STAVGSNMOAIGFAATAVG-----STQVN---LNRGIALGFGSOVLQKDNVNA 287
 DB 401 smaygkakavagalaigaevaagaefdsqagullnrfagyaliks--adksddikag 458
 QY 288 N-VRAVAPDNDPINDRRKATFEKNATDVFSIGNSGNDSTIRKTIINAGASADPDANV 346
 DB 459 dainvftgfdmlycgshlyentlylttsagd-----lkktaavgdgg----- 504
 QY 347 AOLKAVALRANQITFEKGDSDNNRVEKGLGTLTTTGGAQOTSAIDHNHNGVONGDLKV 406
 DB 505 ---knaiaignk-----taskanvalasysalaesgn----- 534
 QY 407 QLAETLTSIKVNTENLTANKEVTVGKTRLTJDKIGFTNDMNGIDESKPYLDKDTGHA 465
 DB 535 --afalgysylvsp--laentliavg-----gyatgsnfsf-----v 567
 QY 466 GGOKITKLKLAGVY-----DDDAATYQQLKKNVQNTAASALQTTVKKVKNDA 514
 DB 568 ggswwslsartvlygsssisdshdflamgvnafignsnssialtgstliakn--a 624
 QY 515 NNSKIITVGNKPKDGTQVNTLKLGENGV-----DVTETNTGT-----VTF 556
 DB 625 kpsdslaigkdsridakdt-----dngvlyltpgyvdecttcafttfdenkndymqama 677

QY 557 GLNQN-----NGLTVG---NSTLND---GLSVKNTNSKQIOVGADITFD 598
 DB 678 gfnakvrygkyketylnsmaigaarsqatlqnsatlayvnahtyvegde-----ad 729
 QY 599 ISNSKPGAGIENTRITRIDGIFANN-----TGSIDANKPRLPTPGIMAGKELTNY 650
 DB 730 pwekgaislptgkligvisvsgyserrlrvnaasgldtd-----avna--qlkcl 780
 QY 651 ---QSAINPATNGQIDFNKRLSTANTKESGSAITKIDLYNLSQVPLRPAOTGRNVK 706
 DB 781 eertgseldllnggygyqy--lsvektlngeagrv-----asql----- 818
 QY 707 KLGE---ILKVGKRTTADLTKNNIGVADSDNLSFLVKLKTTSDDAVNTKLTAS 762
 DB 819 rkgesykyryklyktqillylarkln---gekfdqslidkiskvaveaeysgel--- 871
 QY 763 DKVTVDSGNNTAKLQNDLT-----FSKQNTGATPATNSKTIGVDL-----KFN 807
 DB 872 -ktaselnrvanglnaetvndtgykfnqktlenatnadsekng--glspqvlaqlk 928
 QY 808 DNNGIALDGTYYTKDKVGFRAKODGSLDKSPYLDKDLKVG-----EVEITTMG- 857
 DB 929 annnylndgak--gqdsiafygqaktseannglakqalaifqansaeaislqnsd 986
 QY 858 -----INAGKALITGLSNTLDATNATP-----GHWYOLG 887
 DB 987 tsmtgavaiqgavltvtagpfpalslqgdsstvanaisltsvminglftnfnfsgpsely 1046
 QY 888 FVDSDDTRRAASIDVFNAGNLKNNGDARDFVSTYTPVFINGNATTAATVATDGRASKV 947
 DB 1047 vlsigtgkerkyvna-aggdsqtsleatngslylntmlklagsvntngnann-1 1104
 QY 948 AYDVNDVDTTHLTGADGNKQIGVKTTLTKTDKADKAINFSVNSGDKALINAKDIA 1007
 DB 1105 atdgtlftnngyggqtdlnal--nvaltkl-----lslateeevvsgeavy 1152
 QY 1008 DNLN---TLAGEI-RMTKGADTALOTFOVKYK-----ENGDDNDADITVVGKDAKT 1057
 DB 1153 dalqgkptvsaeanakiqlyvdy-----vkkanspietvpsldmkkkftfyglmkdl 1206
 QY 1058 HQVNTLKLKGNKDIDITNKGIVTFEINTQSG-LKAGNNTTLNNGLSIKRTAGNEQ-- 1114
 DB 1207 egvnsitf-dksqgdlnvgvgrmssagltfkkgdlttngsttttaedltdststnsqtn 1265
 QY 1115 -IQVGADGVKFAKVNNG-----VVGAGIDGTRITRDEIFGAGTNGSLDKSPH 1162
 DB 1266 lvksrdgfv--svkngsdeeklastklsigaenaehvetksjalakadnts-dksait 1321
 QY 1163 LSKRGINNGK-----KIRNIOGELTAONSNDNAVTVGKRIYDL-----KTEL 1203
 DB 1322 lagdaftlagatgtaikltgvaagnltvnskkdavnngqqlrtllgyvdsagkigtetltl 1381
 QY 1204 ENKISSTAKTQNSLHESVADDEOGNNEFTVSNPSSVDTSTSPVTFEAGNGITTVK-N 1262
 DB 1382 seaisdvkqaltatlaykadanngk-----tkkltdgnlftstnldasved 1429
 QY 1263 KGVVRVGI-DQTKGLTTPPKLTVGNNNGKGIYDSQN-GONTITGLSNTLANVTNDKGSVR 1320
 DB 1430 ngvvkftlkdkltgltk-----lateslnasgnliagvtvtygget--egivl 1475
 QY 1321 TTEOGKIIKDEKTRRAASIVDVLSAGFNLQNGCAVDFVSTYTPVDVNRADGNATTAKYTYD 1380
 DB 1476 t-----ksgsngdrtlis-----lsaganaatdgikvsgvka-----gladdavankg 1517
 QY 1381 DTSKTSKVYVNVVDDPTIEVKDKKLGVKTTTLTSTGFGANKPALNSQATGDALVKAASDI 1440
 DB 1518 glklklkainda-----lg--ttldlavt-knpngesirfnpingtepttkda 1561
 QY 1441 VAHLNLTSGDIQTAKGASQANSSAGY--VDADGNKVVITYSDTDNKRYQAKNDGIVDKTREV 1498
 DB 1562 vdklilta--vntgwsakvlglatygdlda-gnkkksnvadgslptspsgd----- 1608

QY 1499 AKDKLVAQAQIPDGTFLAQMNNKSVINKQVNDANKKQGI---NEDNAFVKLEKASDNK 1555
Db 1609 -----vltgrqly-almqkgirygdevspckqctaptlss 1644
QY 1556 TKNAAVTVGDINAVAQPTLTFAGDTGTAKKIGETLTIKGGOTDINKLTDNNIGVACTD 1615
Db 1645 tgggaltantlaggvapagnvatgdlapcpalpenkt-----alvgdhlavp199 1694
QY 1616 GFTVKLAKDLTFLNLSVNNAGGTIKIDDKGVSF-----VSSGQAKAMTPVLSANGL 1664
Db 1695 --slkthgdhnykcttisag---ngvgislgpnlsiemnlvgisnkppekaklaaqegnal 1748
QY 1665 DLGKVYSNVGKGTQDTDAANYQQLNEVRNLLGLGNAGNDNADGNQVNIADIKDPNS-- 1722
Db 1749 -----vltu---kddgnaamv--fneknmlvl-----sdckakpravl 1782
QY 1723 -GSSSRPTVIRKAGTVLGGKNDTEKLGQVGVVDKGNANGLSNVWVKTOKDGSKK 1781
Db 1783 dggngaltlv-----gnaddsq--vlseskkgldgn--dlstlsvltet--- 1824
QY 1782 ALLATVYNAAGQTNVLTNNPEAIDRINEGIRFFHVNDGNQEPVYQGRNGIDSSASGRHS 1841
Db 1825 -----nadgqlckvetstfmdtmd---g1kf----- 1846
QY 1842 VAIGFOAKADGEAAYAIGRQTOAGNOSTAIGDNAQATGDQSIATIGTGNVNVACKHSGAIGD 1901
Db 1847 -----kddgkv1-----nkk1netvelvgdenvt----- 1871
QY 1902 PSTVKADNYSYGVNNNOFTDATQTDVEGVCNNITVTESNSVALGNSNAISACTHAGTOAK 1961
Db 1872 -tsltdgnkvkslnk1a-----idevkiptcdpaqkysdlvimgghagnkv1gvyk 1926
QY 1962 KSDGTAGTTTTAGATGTVKGAGOTAVGAVSVGASGAERRIONVAAGEVSATSTDVNGS 2021
Db 1927 asd-----dp1savnrng 1938
QY 2022 QLYKATQGIAMATNELDRIHONENKANAGISSAMAMASMPQAYIPGRSMVTG1ATIHNG 2081
Db 1939 qlntv1dnvgnfnfngvngf1gdl1resrag1agamataslgnvalpykctlisvg1atfky 1998
QY 2082 OGVAVANGLSKLSDNGQWVFKINGSADTOGHVGAAGVAGCFHF 2122
Db 1999 enava1gm1r1sdngkv1g1r1sgms1c1sngdk1gaamsvg1tf 2039

Search completed: July 30, 2002, 15:35:11
Job time: 466 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:34:52 ; Search time 33.83 Seconds
(without alignments)
1532.105 Million cell updates/sec

Title: US-09-813-214A-9
Perfect score: 10708
Sequence: 1 MNHRYKVFNNKATGTFMAYV.....NCSADTQGHVCAVAGAGHF 2122

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|---------|-------------|--------|-------------------|-------------------|
| 1 | 10697.5 | 99.9 | 2123 | US-08-968-685A-10 | Sequence 10, Appl |
| 2 | 7320 | 68.4 | 2048 | US-09-268-347-48 | Sequence 48, Appl |
| 3 | 6585 | 61.5 | 2314 | US-09-268-347-49 | Sequence 49, Appl |
| 4 | 1185.5 | 11.1 | 2353 | US-09-377-155-33 | Sequence 33, Appl |
| 5 | 1185.5 | 11.1 | 2353 | US-08-913-942-4 | Sequence 4, Appl |
| 6 | 1185.5 | 11.1 | 2353 | US-09-669-974-33 | Sequence 33, Appl |
| 7 | 1177 | 11.0 | 2354 | US-09-268-347-47 | Sequence 47, Appl |
| 8 | 1169.5 | 10.9 | 2411 | US-09-268-347-36 | Sequence 36, Appl |
| 9 | 1018.5 | 9.5 | 1104 | US-09-268-347-38 | Sequence 28, Appl |
| 10 | 1018.5 | 9.5 | 1104 | US-09-268-347-34 | Sequence 34, Appl |
| 11 | 974 | 9.1 | 1004 | US-09-268-347-30 | Sequence 30, Appl |
| 12 | 944.5 | 8.8 | 1002 | US-09-268-347-24 | Sequence 24, Appl |
| 13 | 845 | 7.9 | 1912 | US-08-409-995-4 | Sequence 4, Appl |
| 14 | 845 | 7.9 | 1912 | US-08-685-467-4 | Sequence 4, Appl |
| 15 | 693 | 6.5 | 1094 | US-09-268-347-32 | Sequence 32, Appl |
| 16 | 681 | 6.4 | 1098 | US-08-409-995-2 | Sequence 2, Appl |
| 17 | 681 | 6.4 | 1098 | US-08-685-467-2 | Sequence 2, Appl |
| 18 | 681 | 6.4 | 1098 | US-09-377-155-32 | Sequence 32, Appl |
| 19 | 681 | 6.4 | 1098 | US-08-913-942-2 | Sequence 2, Appl |
| 20 | 681 | 6.4 | 1098 | US-09-669-974-32 | Sequence 32, Appl |
| 21 | 681 | 6.4 | 1098 | US-09-268-347-44 | Sequence 44, Appl |
| 22 | 582.5 | 5.4 | 679 | US-08-913-942-15 | Sequence 15, Appl |
| 23 | 582.5 | 5.4 | 679 | US-09-268-347-26 | Sequence 26, Appl |
| 24 | 542 | 5.1 | 1600 | US-08-617-697-10 | Sequence 10, Appl |
| 25 | 525.5 | 4.9 | 1599 | US-08-617-697-9 | Sequence 9, Appl |
| 26 | 509 | 4.8 | 1536 | US-08-038-682-2 | Sequence 2, Appl |
| 27 | 509 | 4.8 | 1536 | US-08-302-832-2 | Sequence 2, Appl |

| | | | | | | |
|----|-------|-----|------|---|------------------|-------------------|
| 28 | 509 | 4.8 | 1536 | 2 | US-08-530-198-2 | Sequence 2, Appl |
| 29 | 509 | 4.8 | 1536 | 2 | US-08-469-880-2 | Sequence 2, Appl |
| 30 | 509 | 4.8 | 1536 | 2 | US-08-728-470-2 | Sequence 2, Appl |
| 31 | 509 | 4.8 | 1536 | 2 | US-08-617-697-2 | Sequence 2, Appl |
| 32 | 509 | 4.8 | 1536 | 4 | US-08-719-641-2 | Sequence 10, Appl |
| 33 | 509 | 4.7 | 1529 | 4 | US-08-728-470-10 | Sequence 10, Appl |
| 34 | 500 | 4.7 | 1529 | 4 | US-08-719-641-10 | Sequence 10, Appl |
| 35 | 487.5 | 4.6 | 1477 | 1 | US-08-038-682-4 | Sequence 4, Appl |
| 36 | 487.5 | 4.6 | 1477 | 1 | US-08-302-832-4 | Sequence 4, Appl |
| 37 | 487.5 | 4.6 | 1477 | 2 | US-08-530-198-4 | Sequence 4, Appl |
| 38 | 487.5 | 4.6 | 1477 | 2 | US-08-469-880-4 | Sequence 4, Appl |
| 39 | 487.5 | 4.6 | 1477 | 2 | US-08-728-470-4 | Sequence 4, Appl |
| 40 | 487.5 | 4.6 | 1477 | 2 | US-08-617-697-4 | Sequence 4, Appl |
| 41 | 487.5 | 4.6 | 1477 | 4 | US-08-719-641-4 | Sequence 4, Appl |
| 42 | 461 | 4.3 | 1338 | 2 | US-08-728-470-9 | Sequence 9, Appl |
| 43 | 461 | 4.3 | 1338 | 4 | US-08-719-641-9 | Sequence 9, Appl |
| 44 | 458.5 | 4.3 | 1612 | 1 | US-08-169-927-2 | Sequence 2, Appl |
| 45 | 401.5 | 3.7 | 599 | 4 | US-09-377-155-15 | Sequence 15, Appl |

ALIGNMENTS

RESULT 1
US-08-968-685A-10
Sequence 10, Application US/08968685A
Patent No. 6214981
GENERAL INFORMATION:
APPLICANT: TUCKER, KENNETH
APPLICANT: PLOSLA, LAURA
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,685A
FILING DATE: No. 6214981ember 12, 1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2123 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-968-685A-10

Query Match 99.9%; Score 10697.5; DB 4; Length 2123;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2122; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
1 MNHRYKVFNNKATGTFMAYVAKSHSTGGSCATGCGSVRLTSPARIALAVLVIGAT 60
|||||

Db 1 MNHIIYKFNKATGEMAVAEYAKSHSTGGSCATQVGSVRTLSEFARIALAVLIGAT 60
 Qy 61 LNSAAYAGIGISEADGGKGGANARPKSTIAGDIAQALGSOSTAIGDNKIVHNSNNAN 120
 Db 61 LNSAAYAGIGISEADGGKGGANARPKSTIAGDIAQALGSOSTAIGDNKIVHNSNNAN 120
 Qy 121 GAKASGESIAIGDVLASGHASIAIGSDLYLKEETVOOISELPIIRGOKALNDIYL 180
 Db 121 GAKASGESIAIGDVLASGHASIAIGSDLYLKEETVOOISELPIIRGOKALNDIYL 180
 Qy 181 ADNLLOKRRTHAOGHASTAVGAMSTAKGHFSNAFGTRATBEETSYSLAVGLRATAAAS 240
 Db 181 ADNLLOKRRTHAOGHASTAVGAMSTAKGHFSNAFGTRATBEETSYSLAVGLRATAAAS 240
 Qy 241 IAVGNAOAGIAPATVAGSTQVNLNRGIALGFSQVLOKNDVNAANRATAPDDNOPI 300
 Db 241 IAVGNAOAGIAPATVAGSTQVNLNRGIALGFSQVLOKNDVNAANRATAPDDNOPI 300
 Qy 301 DNRYKATFKNGATVFSIGNSNGNDSIRKKIIVGAGSADTDVAVVAVOLKEAVRLANRO 360
 Db 301 DNRYKATFKNGATVFSIGNSNGNDSIRKKIIVGAGSADTDVAVVAVOLKEAVRLANRO 360
 Qy 361 TFKGDSNNRVEKIGKTLITGGAQTSALTQHNIGVONGOLKQOLAEITLSLMTT 420
 Db 361 TFKGDSNNRVEKIGKTLITGGAQTSALTQHNIGVONGOLKQOLAEITLSLMTT 420
 Qy 421 ENLTANEKYVCKRLTPTDKIGFTNDMNGIDESKPYLDKDTGTHAGGOKITKLTAGVVD 480
 Db 421 ENLTANEKYVCKRLTPTDKIGFTNDMNGIDESKPYLDKDTGTHAGGOKITKLTAGVVD 480
 Qy 481 DAATYQOLKVVNOJASALQTFVKKVYKNDANDANSKIITVGNKNNPDGTQVNTLKLKG 540
 Db 481 DAATYQOLKVVNOJASALQTFVKKVYKNDANDANSKIITVGNKNNPDGTQVNTLKLKG 540
 Qy 541 ENGVDYTTENGVTGGLMONNGLTYGNSLTNNGLSVKNTNSNQIOYGAGITPTDLS 600
 Db 541 ENGVDYTTENGVTGGLMONNGLTYGNSLTNNGLSVKNTNSNQIOYGAGITPTDLS 600
 Qy 601 NSKPGAGIENTRITRDGIGFANNTGSLDANKPRLPTGINSAGKELTJVWOSAIPATNG 660
 Db 601 NSKPGAGIENTRITRDGIGFANNTGSLDANKPRLPTGINSAGKELTJVWOSAIPATNG 660
 Qy 661 GOLDFENRISTANTEKSSGAATIKDLYNLSQVPLTFAGDTGPNVTKLCEIILVKGKTT 720
 Db 661 GOLDFENRISTANTEKSSGAATIKDLYNLSQVPLTFAGDTGPNVTKLCEIILVKGKTT 720
 Qy 721 ADDLTNNNGVAVADSTNLSLTVKIAKTSLDLAVNMTKTLTASDKVTVDSGNNTAKLQND 780
 Db 721 ADDLTNNNGVAVADSTNLSLTVKIAKTSLDLAVNMTKTLTASDKVTVDSGNNTAKLQND 780
 Qy 781 LFFSKONTGATPATNSKTI-GVDGLKFTDNGIALDGTYYIRKDVYGFAPKOGSILDKSP 839
 Db 781 LFFSKONTGATPATNSKTI-GVDGLKFTDNGIALDGTYYIRKDVYGFAPKOGSILDKSP 839
 Qy 840 YLDKDKLAKGEVEITTINGINAGKALITGLSNTLDTATNAATGHVTOGLIVSDTKTRAAS 899
 Db 840 YLDKDKLAKGEVEITTINGINAGKALITGLSNTLDTATNAATGHVTOGLIVSDTKTRAAS 899
 Qy 900 IGDVLNAGNKLKNGDAKFEVSTYDTPVDFINGNATTAKTATYVGRKSKAYAVVNVGCTI 959
 Db 900 IGDVLNAGNKLKNGDAKFEVSTYDTPVDFINGNATTAKTATYVGRKSKAYAVVNVGCTI 959
 Qy 960 LFGADGNKNOIGVKTTLTKTDKADKALINFVNSGDDKALINADIDNLTLTLAGETRN 1019
 Db 960 LFGADGNKNOIGVKTTLTKTDKADKALINFVNSGDDKALINADIDNLTLTLAGETRN 1019
 Qy 1020 TFGTADTALQOTVQVKKVKNEDDDADDTIYVGDKAKTNQVNTLKLKKGNGIDTQNDK 1079
 Db 1020 TFGTADTALQOTVQVKKVKNEDDDADDTIYVGDKAKTNQVNTLKLKKGNGIDTQNDK 1079
 Qy 1080 TWTFGINOSGLKAGNNTLNNGLSIKNTAGNEOIOYGADSVKRAKAVNNGVAGIGICT 1139
 Db 1080 TWTFGINOSGLKAGNNTLNNGLSIKNTAGNEOIOYGADSVKRAKAVNNGVAGIGICT 1139
 Db 1081 TWTFGINTOSGLKAGNNTLNNGLSIKNTAGNEOIOYGADSVKRAKAVNNGVAGIGICT 1140

Qy 1140 TRTRDEIGFAGTNGSLDKSPHLKSGKOGINAGKKTINIOSGELAONSNDATGKTYDL 1199
 Db 1141 TRTRDEIGFAGTNGSLDKSPHLKSGKOGINAGKKTINIOSGELAONSNDATGKTYDL 1200
 Qy 1200 KTELENKISSTAKTAQNSLHFSVADDEGNNFTVSNPSSYDTSKTSVITFAGENGITT 1259
 Db 1201 KTELENKISSTAKTAQNSLHFSVADDEGNNFTVSNPSSYDTSKTSVITFAGENGITT 1260
 Qy 1260 KYNKGVAVGIDOTKGLTTPKLTJGNNNGKGIIVDSQNGOVTITGLSNTLANVTNDKGSV 1319
 Db 1261 KYNKGVAVGIDOTKGLTTPKLTJGNNNGKGIIVDSQNGOVTITGLSNTLANVTNDKGSV 1320
 Qy 1320 RTTEBOKIIRKEDKTRASIVDVSAGNLTGNGEAVDFVSTYDTPVADGNATTAATVY 1379
 Db 1321 RTTEBOKIIRKEDKTRASIVDVSAGNLTGNGEAVDFVSTYDTPVADGNATTAATVY 1380
 Qy 1380 DTSKTSKRVYDVANVDPTTIEVKDKKLGKTTTLTSTGTGANKFALSNQATGALVRASD 1439
 Db 1381 DTSKTSKRVYDVANVDPTTIEVKDKKLGKTTTLTSTGTGANKFALSNQATGALVRASD 1440
 Qy 1440 IVAHLNTLSGDIQAKGASQANSSAGYVDADGNKYIYDSTDNKTYYQAKNDGTVDKTEVA 1499
 Db 1441 IVAHLNTLSGDIQAKGASQANSSAGYVDADGNKYIYDSTDNKTYYQAKNDGTVDKTEVA 1500
 Qy 1500 KDKLVAQAOTPDGTLAONNVKSVINKBEQVNDANKKOGINEBNAPVKGLEKASDNKTKNA 1559
 Db 1501 KDKLVAQAOTPDGTLAONNVKSVINKBEQVNDANKKOGINEBNAPVKGLEKASDNKTKNA 1560
 Qy 1560 AVTVGDLNAVAOTPLTFAGDTGTAKKLGELTLTKGQOTDNTKLTJNNIGVAVGDTGFTV 1619
 Db 1561 AVTVGDLNAVAOTPLTFAGDTGTAKKLGELTLTKGQOTDNTKLTJNNIGVAVGDTGFTV 1620
 Qy 1620 KLANDLTMLNSVNAAGRTITDKKVSFVDSQGAANPVLISANGLDGGKVISVWGKTK 1679
 Db 1621 KLANDLTMLNSVNAAGRTITDKKVSFVDSQGAANPVLISANGLDGGKVISVWGKTK 1680
 Qy 1680 DTDAAVQOLNEVRNLGLGAGNADNDGNOVNIADIKKPDNSSSSNRPVIRAGTYLGG 1739
 Db 1681 DTDAAVQOLNEVRNLGLGAGNADNDGNOVNIADIKKPDNSSSSNRPVIRAGTYLGG 1740
 Qy 1740 KGNNDTEKATGVQVGVYDKDGNANGDLSNVVYKTDGSKKALLATYNAAGOTNYLTNN 1799
 Db 1741 KGNNDTEKATGVQVGVYDKDGNANGDLSNVVYKTDGSKKALLATYNAAGOTNYLTNN 1800
 Qy 1800 PAEALDRINEGIRFFHVNDGNDEPVVQGRNGIDSSASGRHSAIGFOAKADEAAVAIG 1859
 Db 1801 PAEALDRINEGIRFFHVNDGNDEPVVQGRNGIDSSASGRHSAIGFOAKADEAAVAIG 1860
 Qy 1860 ROTAGNOSIAGDMAOTGPOSIAIGTGNVYVAKRHSGAIGDPSTYKADNSYSVGNNOF 1919
 Db 1861 ROTAGNOSIAGDMAOTGPOSIAIGTGNVYVAKRHSGAIGDPSTYKADNSYSVGNNOF 1920
 Qy 1920 TDAOTQDVEGVGNNTIYTESNSVALGSNSAISACTHAGTQAKKSDGAGTTTATAGATV 1979
 Db 1921 TDAOTQDVEGVGNNTIYTESNSVALGSNSAISACTHAGTQAKKSDGAGTTTATAGATV 1980
 Qy 1980 KGFAGQTAHVAVSGASGAERIRIONVAAGEVSASTVAVNGSOLYKATOGIAATNELDH 2039
 Db 1981 KGFAGQTAHVAVSGASGAERIRIONVAAGEVSASTVAVNGSOLYKATOGIAATNELDH 2040
 Qy 2040 RIHONENKANAGISSAAMAMAMPQATIPGRSMYTGATINNGGCAVAVGSLKSDNGOW 2099
 Db 2041 RIHONENKANAGISSAAMAMAMPQATIPGRSMYTGATINNGGCAVAVGSLKSDNGOW 2100
 Qy 2100 FKINGSADTOGHVGAAGAGFHF 2122
 Db 2101 FKINGSADTOGHVGAAGAGFHF 2123

RESULT 2
 US-09-268-347-48
 ; Sequence 48, Application US/09268347

Patent No. 6335182
 GENERAL INFORMATION:
 APPLICANT: Loosmore, Sheena M.
 TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
 FILE REFERENCE: 1038-860
 CURRENT APPLICATION NUMBER: US/09/268,347
 NUMBER OF SEQ ID NOS: 54
 CURRENT FILING DATE: 1999-03-16
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO: 48
 LENGTH: 2048
 TYPE: PRN
 ORGANISM: Haemophilus influenzae
 US-09-268-347-48

Query Match 68.4%; Score 7320; DB 4; Length 2048;
 Best Local Similarity 68.1%; Pred. No. 0;
 Matches 1552; Conservative 117; Mismatches 221; Indels 390; Gaps 24;

OY 1 MNHYKVFENKATGTPMAVAEYAKSHSTGGSCATGQVGSRTTSPARIALAVLVIGAT 60
 DB 1 MNHYKVFENKATGTPMAVAEYAKSHSTGGSCATGQVGSRTTSPARIALAVLVIGAT 60
 OY 61 LNSGAYV-----GIGISEADGKGGANRGKSTAGIDIAQALGOSIAIGDKIVH 112
 DB 61 LNSGAYVAKKTKTHAIGEQNPPRRSGTAKADDBAIAIGENANAQGOAIAIGSSNKTIV 120
 OY 113 NSNNANNAKAGASNESIAIGDVLASGHASIAIGSDLYL-----KKEITVOQISELL 165
 DB 121 NSSLDKICTDATGQESIAIGDVKASGDASIAIGSDLLHLDHGNKHPKGLINDL- 179
 OY 166 PLIRQAKLNDIYQALDNLQKRYTHAOGHASTAVAMSTAKHFSNAFTKATBETTY 225
 DB 180 --INGHAULKEIRSSKMDV- KYRRTTASGHASTAVAMSAQGHFSAFTKATASAY 236
 OY 226 SLAVGLTPTAKAASIAVGSNAQATGFAATVAGSTOVNLNGLALGSGOVLQKNDVYN 285
 DB 237 SLAVGLATAGQSTIAIGSDATSSSLGALGAGTRAOLOGSLALGSGSVYVOSDN--- 293
 OY 286 AANVR-AVAPDNDOPIDNRKATEFKNGATDVFSIGNSNGNOSIRKIIIVAGSADTAV 344
 DB 294 --NSRPATTP-NTQALDPRKFOATNNTKAGPL-----SIGNSIKRKIIIVAGVAKTDAV 345
 OY 345 NVAOLKEAVRLA-NRQITFEKDDSNRRVEKGLGTLITGGAQISALTLDHITGVQND- 402
 DB 346 NVAOLEAVVAKKERIRITFOGDDNSTDVKIGLDNTLITKGAEETNALDNNIGVYKEDN 405
 OY 403 -GLKQALETLTSKMTTEMLTANEKVTGK----- 433
 DB 406 SGLKVKLAKTLNLTLEVTTLNATTYKVGSSSSTTAEILLSLSLFTQPTNGSOSTSKT 465
 OY 434 -----TRLTTRKIGFTNDNGIDSEKPYLDK-----D 460
 DB 466 VYGVNGVFTNNAETTAIGTRITRIRDKIGFARDDVDEKQAPYDKQKLVGSAVITID 525
 OY 461 TGIHAGGQIKTKLAGVDDDAATYGQL----- 489
 DB 526 NGIDAGNKKISMLAGSSANDAVITIEQLKAAKPTLNAAGASIVPTETISVDKSGSVTAP 585
 OY 490 -----KVQNTAESALQPTTVKV 507
 DB 586 TYNIGVKTTELSNDSGTSPKFSVSGSTNNSLVTAHSLASYLEVVKRIVDSALQSTTYAE- 644
 OY 508 DKNGDANDSKLITVGNKNNKPDGTOVNTLKLKGENGVDTTETNGTTFGLNONGTLVVG 567
 DB 645 -----EDDDANAIYAKDPTTKNAGAVSILKLKGNGLVATRKQDQTYVEGLSDGGLTIG 700
 OY 568 NSTLNDGILSVANTNSKQIOYAGDGTFTDISNKKPAGIENRTIRITRDGIGFANNTGS 627
 DB 701 KSTLNDGILTVKDT--NRDIOVGANGIKFTYVNGSNPFGTIGLANTARITR----- 747
 OY 628 LDANKPRLPTGINGAGKELTNVQSAINDPATNGQLDFNNRLSTANTKSGSAATIKDLY 687

DB 748 ----- 747
 OY 688 NLSQVPLTFPAGDTGPVNTKGLLEIKVKGKITTADDLTNNIGVADSTDNSLITVLAKT 747
 DB 748 ----- 747
 OY 748 LSLDLAVNFKTLASDKVTVDSGNNTAKLQNGDLFESKQNTGATPATNSKTIGVGLKFT 807
 DB 748 ----- 747
 OY 808 DNNGLADGTTTYITKDKVGFAPKQDSLDKSPYLDKDKLVGEVEITTTNGINAGKAITG 867
 DB 748 -----DKIGASGSDAVDTKRPYLDQDKLQYGVKTLTNGINAGKAITG 792
 OY 868 LSNLTDAINATGHTVTOIGTVDSTKTRASTIGVLANAGFNILKNNGAKDEYSTVDYD 927
 DB 793 LSPTLPSIADQSSRNIELENTIDQDKSNMAASINDILNTGFNLKNNNPLDFVSTYDIYD 852
 OY 928 FINGNATTAATVYD--GKASKVAVYDVNDGTTIHLGADGNKQOIGVKTTLTKTPAKGD 985
 DB 853 FANGNATTAATVYDTPANKSKVYDVAVDDTTIHLGTIDDNK-KLGVKTKTLKTSANGN 911
 OY 986 KAINFVSNGDDKALINAKDIADNLNTLAGEIRNTGTADTALQTPQVKKVKEGDDND 1045
 DB 912 EATNFVNSSDEDAVNAKDIADNLNTLAGEIRNTGTADTALQTPQVKKVKEGDDND-D 970
 OY 1046 ADITVYKDAKTNQVNTLKLKGNGLDICTNKQGYTFEGINTOSGLKAGNNTTLNNGLS 1105
 DB 971 ANATVYQKANNQVNTLTLKGEINIKTKDKNGYTFEGINTOSGLKAG-KSTLNDGGLS 1029
 OY 1106 IKNTAGNEOIQVGDGKFPKV--NNGVYAGIDGTRITRITDEIGTFAGTNGSLDKSPHLS 1164
 DB 1030 IKNPTGSEQIQVGDGKFPKVNNGVYAGIDGTRITRITDEIGTFAGTNGSLDKSPHLS 1089
 OY 1165 KDGINAGKKITNIOSEIQAONSNDATVYTGK IYDLKTELENKISSTAKTAQNSLHEFSYA 1224
 DB 1090 KDGINAGKKITNIOSEIQAONSHDVAVTGKIYDLKTELENKISSTAKTAQNSLHEFSYA 1149
 OY 1225 DEQGNNTVSNPSSVYDTSKSDVITTFAGENGITTKVKNKVAVYGDQTKGLTTPLYTG 1284
 DB 1150 DEQGNNTVSNPSSVYDTSKSDVITTFAGENGITTKVKNKVAVYGDQTKGLTTPLYTG 1209
 OY 1285 NNNKGIVIDSONQONTITGSLNTLAVTNDKGSVTRTEQGIITKEDKTRASIVDVS 1344
 DB 1210 NNNKGIVIDSONQONTITGSLNTLAVTNDKGSVTRTEQGIITKEDKTRASIVDVS 1269
 OY 1345 AGFNLOGNGEAVDFVSTYDVFADGNATTAATYDTSKTSKVYVDVNVDDTTIEVKDK 1404
 DB 1270 AGFNLOGNGEAVDFVSTYDVFADGNATTAATYDTSKTSKVYVDVNVDDTTIEVKDK 1328
 OY 1405 KLGKVTTLTSTGTGANKKALSNQATGDALVYKASDIYAH-LNTLSGDICTAKASQANS 1463
 DB 1329 KLGKVTTLTSTGTGANKKALSNQATGDALVYKASDIYAH-LNTLSGDICTAKASQANS 1388
 OY 1464 AGVYDADGNK-VIYDSTDNRYOAKNDGTVDKTEVAKDLVLAQOTPDSTLQOMNVKSV 1522
 DB 1389 AGVYDADGNKVIYDSTDNRYOAKNDGTVDKTEVAKDLVLAQOTPDSTLQOMNVKSV 1448
 OY 1523 INKEQVNDANKKQGINEDNFKVGLERAASDNKTKNAAVVGDILNVAOQPTLPFAGDTGT 1582
 DB 1449 INKEQVNDANKKQGINEDNFKVGLERAASDNKTKNAAVVGDILNVAOQPTLPFAGDTGT 1508
 OY 1583 TARKLGETLITKGGCTTNKLTNNICVAGTDCGFYVYKAKDLTNLNSVAGGTIKDDKG 1642
 DB 1509 TARKLGETLITKGGCTTNKLTNNICVAGTDCGFYVYKAKDLTNLNSVAGGTIKDDKG 1568
 OY 1643 VSFVDSGQAKANPVPVLSANGLOLGVKISVNGGCTDTPAANVOOLNEVRNLGLGNAG 1702
 DB 1569 VSFVDSGQAKANPVPVLSANGLOLGVKISVNGGCTDTPAANVOOLNEVRNLGLGNAG 1628
 OY 1703 NDNDAGNQVNTADIKKDPNSGSSSNRTVIRAGTVLGGKGNNDKELATGQVGVYVDKGN 1762

Db 1629 NDNADGNQVNIADIKDPNSGSSNRTVIKAGTVLGGKGNNDTEKLATGIGVGVDCDN 1688
 QY 1763 ANGDLNVMVWVKQKDGSKKLLATYNAGOTNLTNNPAEIDRIEQQIRPHVNDGQ 1822
 Db 1689 ANDDLNVWVWVKQKDGSKKLLATYNAGOTNLTNNPAEIDRIEQQIRPHVNDGQ 1748
 QY 1823 EPVVOGNGIDSSASGKSHVAIGFOAKKADGEAAVAIGROTOAGNOSIAGDNOATGDS 1882
 Db 1749 EPVVOGNGIDSSASGKSHVAIGFOAKKADGEAAVAIGROTOAGNOSIAGDNOATGDS 1808
 QY 1883 IAGTGNVAGKSHGAIIGDPSTYKADNSYSVGNNOFTDATTOTDVEGVNNTVTESSNV 1942
 Db 1809 IAGTGNVAGKSHGAIIGDPSTYKADNSYSVGNNOFTDATTOTDVEGVNNTVTESSNV 1868
 QY 1943 ALNSNSAISGTHAGTQAKSDGTAGTTAGTGTGKGPAGOTAVGAASVAGSGAKERI 2002
 Db 1869 ALNSNSAISGTHAGTQAKSDGTAGTTAGTGTGKGPAGOTAVGAASVAGSGAKERI 1928
 QY 2003 QNVAGEVSATSTDAVNGSOLYKATOGIANATVELDHRITHONENKANAGISSAMAMASMP 2062
 Db 1929 QNVAGEVSATSTDAVNGSOLYKATOGIANATVELDHRITHONENKANAGISSAMAMASMP 1988
 QY 2063 QAVIPGRSMVTGGIATHNGOGAAVGLSKLSDNGQVVEFKINSADTQGHVGAAGAFHF 2122
 Db 1989 QAVIPGRSMVTGGIATHNGOGAAVGLSKLSDNGQVVEFKINSADTQGHVGAAGAFHF 2048

RESULT 3

US-09-268-347-49
 ; Sequence 49, Application US/09268347
 ; Patent No. 6335182
 ; GENERAL INFORMATION:
 ; Applicant: Leosmore, Sheena M.
 ; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
 ; FILE REFERENCE: 1038-860
 ; CURRENT APPLICATION NUMBER: US/09/268,347
 ; CURRENT FILING DATE: 1999-03-16
 ; NUMBER OF SEQ ID NOS: 54
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 49
 ; LENGTH: 2314
 ; TYPE: PRP
 ; ORGANISM: Haemophilus influenzae
 ; US-09-268-347-49

Query Match 61.5%; Score 6585; DB 4; Length 2314;

Best Local Similarity 57.1%; Pred. No. 0;

Matches 1460; Conservative 137; Mismatches 280; Indels 682; Gaps 40;

QY 1 MNHIVYVFNKATGTFMAVAEYAKSHSTGGSCATGQVGSVFTLSFARIALAVIGAT 60
 Db 1 MNHIVYVFNKATGTFMAVAEYAKSHSTGGSCATGQVGSVFTLSFARIALAVIGAT 60
 QY 61 LNSGAVAGIGISEADGKG--GANARGDSIAIGDIAQALGOSISAIADNR-IVHNSNN 116
 Db 61 LNSGAVAGIGISEADGKG--GANARGDSIAIGDIAQALGOSISAIADNR-IVHNSNN 116
 QY 117 NANIGAKAGNSIAIGGVYLASGHAASIAIGSDDLILKKEVYQ--ISELLPIIRGQKA 173
 Db 121 NOKAGSHAGKRSIAIGGVYLASGHAASIAIGSDDLILKKEVYQ--ISELLPIIRGQKA 180
 QY 174 LNDIYVADPTN-LQKRYRTHAGHASTAVGAMSYAKGHSNAGFRATPEGTSTAVGLT 232
 Db 181 LR---QIRSGNSQKRYRTHAGHASTAVGAMSYAKGHSNAGFRATPEGTSTAVGLT 237
 QY 233 AFAKAASSTAVGMAOAGFAATAVGSGTQVNLNRCIALGFGSOYLQKDNVNAANVAY 292
 Db 238 AKAEEKYTTAIGSNMOAIVYGALALCADTRVDLDYGIAGYGSQIL---NNNNNNNNKAY 294
 QY 293 ADDNDPIONRRKATPKKATDVFSIGNSNGDSIRKRIINAGSADTDVAVNAOLKEA 352
 Db 295 VPEGNSNIKSKAT-GNG---LFSIGSS---TIKRRIINAGYEDTDVAVNAOLKAV 346

QY 333 VRLANQITFEKGDSSNNREKGLITLTITGG-AQTSALTD-HNIGV-QNGDLKYOQA 409
 Db 347 ENLARKQITFEKGDSSNNREKGLITLTITGG-AQTSALTD-HNIGV-QNGDLKYOQA 406
 QY 410 ETLTSLKAVTTEENLANKKAVTVGK----- 433
 Db 407 KLSGLEIVSTKRLASKEVYVSGSNNFAELQSGGLTPTTNASTDYKVTGIGLAKFTD 466
 QY 434 -----TLTLDKIGFTNDMGIDSKPYLDKYGIGHAGGKITKLAVVDDAAT 484
 Db 467 NSMTALEDTTTRITTKKIGFNSKAGTVDENKPYLDK----- 502
 QY 485 YQGLKKNVQTESALQITFVKKVDKNGDANDSKITTYGKNNKPDGQVNTLKIKGENGV 544
 Db 503 -----KTK----- 505
 QY 545 DYTTEGTGTVTEGLNONGLTIVGNSLTNNDGLSVKNT--NSNKOIOVAGADGTFPTDISNS 602
 Db 506 -----VGNSTLNNGLTJVNTTIGSSNKQIOVAGADGIRKADYVNV 544
 QY 603 KPGAGIENTTTRTRDGI GFANNTGSLDANKPRL-----TPTGINAGKELTN 649
 Db 545 VSNNAKFEFTTRTEBEIEFADADGKVDKSPYLDKQLOVGVKRTKDSGINAGDQKISN 604
 QY 650 VQSA-----INPATNGOLDFEMNRLSTANTKSSGAATIKDLYNLSQV-- 693
 Db 605 VKDATDITDAVTVYQKOVQODADGALQSFIR-----DEKGEFTISNYSNGNTPNT 658
 QY 694 ---LTFAGDTPPNVTKKLGELIKVGGKTADDELTKNNIGVAASTNSLTVKLAKTLSD 750
 Db 659 FETITFAGENGSI-----SNDLAKGVK-----G 684
 QY 751 LDAVNTKTLTASDKVTVDSGNNTAKLONGLTFESKONTGATPATNSKTIGVDGLKFTDNN 810
 Db 685 IDPIN-----GLTTPKL-----TVGSD----- 701
 QY 811 GIALDGTITRKDKVFAKQDGLSKSPYLDKKLAVGEVEITNGINAGKAITGLSN 870
 Db 702 ---KDKTQLV-----LEOVASG-NGTKNIIRKVS 728
 QY 871 TLTDATNATGTVTLQ--IVDSTDKTPRASIGDVLNAGFNLLKNNGDAKDFVSTYDVF 928
 Db 729 TLPSITNNGVTRTEEGNTIISDEDKSAASIGDILMTGFLKNNSSVGVSTYNTVDF 788
 QY 929 INGNATTAKVYD--GKASKVAVYVNVDTGTHLTFGADGKNQOIGVTTTLTKTDAGDK 986
 Db 759 IDGNATTAKVYDETNTQSVTVYVNVDEKTEIETLGGNGKTKNGIGVKTTLTTNANG-K 847
 QY 967 AINFSVNSGDDKALINKDIDNLTLAGELRNKGTADTALQFFOYKAKKENGDDONDA 1046
 Db 848 AINFSST--DNDALYNKADIAENNTLAKELHTTKGTADTALQFFOYKAKKENGDDONDA 900
 QY 1047 DLTIVGKDAKTN--OVNTLKLKNGKGLDITQNKDGTVTFGINTOSGLAKAGNNTTLNNGGL 1104
 Db 901 ETLTSLKAVTTEENLANKKAVTVGK----- 433
 QY 1105 SIKNTAGNEQIOVAGADGKFAKVNNGVYAGCIDGTTTRTDEIGFAGTNSGLSKRPHLS 1164
 Db 961 SIKNPASNEQIOVAGADGKFAKVDKNGSSTGIDTSRTITDOIGFTGANSLSLTTKPHLT 1020
 QY 1165 KD-----GINAGKRTITNIOGSEIAJONSNAVAGKLYDTELENNKISSTAK 1212
 Db 1021 KDKLKGVEITETNGINAGKRTITNIOGSEIAJONSNAVAGKLYDTELENNKISSTAK 1080
 QY 1213 TAONSLHEFVADBOGNNFTVSNPYSSYDTSKTSADVITTFAGENGITTVKVKGVAVGIDQ 1272
 Db 1061 TAONSLHEFVADBOGNNFTVSNPYSSYDTSKTSADVITTFAGENGITTVKVKGVAVGIDQ 1140
 QY 1273 EKGITLTPRLTYGNNNGGIYVDSONGONTITGLSNTLANVTNDKGSVRTTEOGKIIDED 1332
 Db 1141 EKGITLTPRLTYGNNNGGIYVDSONGONTITGLSNTLANVTND-GAGHALSOG-LANDTD 1198

| | | | |
|----|------|---|------|
| Qy | 1333 | KTRAASTVYDVLASGFENLQNGEAVDEVSVDYVDPVNFADGNAAFTAKTYDTSKTSVVYDV | 1392 |
| Db | 1199 | KTRAASTGVDVLNAGFNLQNGEAVDEVSVDYDVEIDGNAAFTAKTYDTSKTSKVYDV | 1258 |
| Qy | 1393 | NVDDTTIEV-KDKLGVKTTTLTSTGTGAN--KFLASNATDAILVAKSDIYAHILNTLSG | 1449 |
| Db | 1259 | NVDKTTIEVTSDKLLGVKTTTLTKTSANGNAIKF---SAADDAILVAKSDIATHLNTLSG | 1315 |
| Qy | 1450 | DIQIAKASQANSAGCYVDADGNKYIYDSTDNKKYQAKNDGVYDKTEVAKDKLVAQAQT | 1509 |
| Db | 1316 | DIQIAKASQANSASSASVYDADGNKYIYDSTDKKYYOVNKGQVDKNKEVAKDKLVAQAQT | 1375 |
| Qy | 1510 | PDGLAAMNKSVMKNOVNDANKKOGIMEDNAFVYGLEKASDUKTKTAAVTVGDLNAV | 1569 |
| Db | 1376 | PDGLAAMNKSVMYVEQVNDANKKOGIMEDNAFIKGLENAKDKTKTAAVTVGDLNAV | 1435 |
| Qy | 1570 | AQPLTFACDGTGTAKKLGETTLTKGGQTDVTKLTDNNIGVAGVDGFTVLAKDLTULN | 1629 |
| Db | 1436 | AQPLTFACDGTGTAKKLGETTLTKGGQTDVTKLTDNNIGVAGVDGFTVLAKDLTULN | 1495 |
| Qy | 1630 | SVNAGTKIDDKGVSPYDSSGQAKANTPYLSANGIDL----- | 1666 |
| Db | 1496 | SVNAGTKRIDEKGISFYDANGQAKANTPYLSANGIDLGGKRISNIGAVDDMDAVNFKQF | 1555 |
| Qy | 1667 | ----- | 1666 |
| Db | 1556 | NEVAKTVNNINNOSSNGSLPFVYTVDANGKPIINGTDGKPOKAIKGADKXYHANANGVPV | 1615 |
| Qy | 1667 | ----- | 1666 |
| Db | 1616 | DKDKPRTTDAADKLANLAHAKPLDAGHQYVYASLGNSDAIYTLTNKSTLPLDTPPNTGNA | 1675 |
| Qy | 1667 | ----- | 1666 |
| Db | 1676 | NAGQAQSLPSLSAAQOSMAASVADVLTNWFNLQTDHNQYDFFKADYDVPFNPGTADTTS | 1735 |
| Qy | 1667 | ----- | 1666 |
| Db | 1736 | VRSDAGTMSNITVNTALATDDGNVLIKAKDGKFYKADDLMPNGSLKAGKSADAKPT | 1795 |
| Qy | 1667 | ----- | 1666 |
| Db | 1796 | GLSLVYNRAGKSTGDVAVALNNLSKAVEFSKDGTTTTVSSDGISIQGKDMSITLSKDG | 1855 |
| Qy | 1667 | ---GGKVISNVGKGTKDTPDANAOULENERBNLLGLGNAGNDADNQVNIADIKKDPNSG | 1723 |
| Db | 1856 | LVNKGKVISNVGKGTKOTPDANAOQLEVRNILLGNGNNDPADNQVNIADIKKDPNSG | 1915 |
| Qy | 1724 | SSSRRTVYIKAGTYLGGKGNNDTEKLTATGGVQYGVDRKGNANGDLSNVWKTQKDGSKAL | 1783 |
| Db | 1916 | SSSRRTVYIKAGTYLGGKGNNDTEKLTATGGVQYGVDRKGNANGDLSNVWKTQKDGSKAL | 1975 |
| Qy | 1784 | LATYNAAGOTNYLTNNPAEADIRINOGIRFPHVYDNGNEPVPVORNCIDSSASGKSHVA | 1843 |
| Db | 1976 | LATYNAAGOTNYLTNNPAEADIRINOGIRFPHVYDNGNEPVPVORNCIDSSASGKSHVA | 2035 |
| Qy | 1844 | IGFOAKADGEAAVAALIGKOTQAGNQSIAIDNDAQTDGDSIAIGTGNVYAKHSGAIGDPS | 1903 |
| Db | 2036 | IGFOAKADGEAAVAALIGKOTQAGNQSIAIDNDAQTDGDSIAIGTGNVYAKHSGAIGDPS | 2095 |
| Qy | 1904 | TVKADNSTSVGNNOFTDATQDVFEGVGNNTVYTESNSVALGNSAISAQTHAGTQAKRS | 1963 |
| Db | 2096 | TVKADNSTSVGNNOFTDATQDVFEGVGNNTVYTESNSVALGNSAISAQTHAGTQAKRS | 2155 |
| Qy | 1964 | DGTAGTTTGATGATGYGAFPGQFAVAVSAGSAGERRIQNVAABEVSATSTDVANGSQL | 2023 |
| Db | 2156 | DGTAGTTTGATGATGYGAFPGQFAVAVSAGSAGERRIQNVAABEVSATSTDVANGSQL | 2215 |
| Qy | 2024 | YKATQGIATANTNELDHRIHQNEKNKAAGISSAMAMASMPQAITPGRSMVTGGIATFHNGG | 2083 |
| Db | 2216 | YKATQGIATANTNELDHRIHQNEKNKAAGISSAMAMASMPQAITPGRSMVTGGIATFHNGG | 2275 |
| Qy | 2084 | AVAVGLSKLSDNGCOWMFKINGSDADTQGHVGAAGAFHE | 2122 |

Db 2216 AVAVGLSKLSDNGQWFKINGSDADTGVHGAVAGGFHF 2314

RESULT 4

US-09-377-155-33

Sequence 33, Application US/09377155

Patent No. 6197312

GENERAL INFORMATION: . . .

APPLICANT: PEAK, Ian Richard Anselm

APPLICANT: JENNINGS, Michael Paul

APPLICANT: MOXON, E. Richard

TITLE OF INVENTION: NOVEL SURFACE ANTIGEN

FILE REFERENCE: 065064/0128

CURRENT APPLICATION NUMBER: US/09/377.155

CURRENT FILING DATE: 1999-08-19

PRIOR APPLICATION NUMBER: PCF/AU98/01031

PRIOR FILING DATE: 1998-12-14

PRIOR APPLICATION NUMBER: GB 9726398.2

PRIOR FILING DATE: 1997-12-12

NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 33

LENGTH: 2353

TYPE: PRF

ORGANISM: Haemophilus influenzae

US-09-377-155-33

| Query Match | Local Similarity | 11.1% | Score 1185.5 | DB 4 | Length 2353 |
|-----------------------|------------------|--|-------------------|----------|-------------|
| Best Local Similarity | 22.7% | | Pred. No. 8.3e-61 | | |
| Matches 613 | Conservative 330 | Mismatches 832 | Indels 925 | Gaps 116 | |
| QY | 1 | MNLIKYIFNKATGTEPMAYAEYAKSHSTGGSCATGYGVSVRTLSPFARIAALAVLIGAT | 60 | | |
| Db | 1 | MNKFNFYIMVMVMTQWVWVSELTFRHTK--RASATVETAVLATLTLFAFYQANAT----- | 52 | | |
| QY | 61 | LNGSAVAGISISDADGGKAGNARGDKSLAIGDIAQ-----ALGSOSIALGDNKIYVNSN | 115 | | |
| Db | 53 | -----DEDELDPPVTRAPVLSPFSHSDKEGTEKEWENS | 87 | | |
| QY | 116 | NNANIGAKAGNESIAIGGDVLASGHASIASDGLYKKEFTVOQISBLPIIRGOKALN | 175 | | |
| Db | 68 | WGIFYFDKG-----VLKAGLITLKAG-DMLKTKQMTDE----- | 119 | | |
| QY | 176 | DIYOLADTNLOKTYRTHAOGHASTAVAGANSYAKGHFSNAFSTRATAREGYSLAYGLT-AT | 234 | | |
| Db | 120 | -----STNASSF----- | 138 | | |
| QY | 235 | AKAASSIYAVSNNOAIGFATAVG-----GSTOVNLRGIALGFGSOYLQKNDVNAAN | 288 | | |
| Db | 139 | SVATREKISFSPANDKVDITSDANGKLAKTNGVNHNL-----GLDSTLPDAVTNIGVUS | 193 | | |
| QY | 289 | VRAAYAPDDNQPINRRKATFEKNGATDVSISIG-NSNGNDSIRRKIIYNAGSADPTDAVNA | 347 | | |
| Db | 194 | SSSPTPMD--VEKTRAAFTVK-----DVLNAGNMING-----AKAGAGNV-- | 230 | | |
| QY | 348 | QKRAAVL--ANQOTFFKGDDSN-----RVEGGLCKTLTITTGCAQTSALDHNIGVQNG | 401 | | |
| Db | 231 | -----ESVDLVASVYNNVFEITGDCKRTLDDVLLTAENCKTEVEKTEPRTS-----VIREK | 279 | | |
| QY | 402 | DGLKVLQALFELTSLKVVTEENTLANEKVYVCGKTRLTQKIGFTDMNGIDESKPYLDKPT | 461 | | |
| Db | 280 | DG-----KLFTEKENNDNRKYT--SNATLN--TDEGMDLYAKAVID-- | 318 | | |
| QY | 462 | GIHAGGOKITKLTAAGVDDAATYQOLKKVNOFAESALQETTVKRVKNGN-----DA-- | 514 | | |
| Db | 319 | AVNAGGRVKTTLANGONGDPATVA--SGTNVTFESGGGTASTAYTKDPTNGNGITVYKIDAK | 376 | | |
| QY | 515 | -----NSKLIITYGKNNKPDGTYVNTL-----KLKGNGCYVTI----- | 548 | | |
| Db | 377 | VGDLGKFDSDSKIVADTTLATLVYTGKVAIEIAKEDKKLVLAAGGLVYALGNLSMKAKAEA | 436 | | |

QY 549 EBNGTVEGLNONNGITVGNSTLNDGLSVK-----NTNSKQIOVGADGDTFTEDISNKP 604
 Db 437 DTDGALD-GISKDOEKAGEITVFRAGKNLKYKOGAGANTYSLODALTGLTSTLGGTN 495
 QY 605 GAGINENTRITRDGIGF--ANNITSLDANKPRLPTGAINAGKELTNVOSAINPATNGO 662
 Db 496 G-GNDAKTVINKDGLITFPAGNGGTGTNTISVTKDGIKAGKALTNVAGSLK-AVDAN 553
 QY 663 LDFMNRLLST-----ANTE--KSSSAITKDLNLSOVPPLTFAG- 698
 Db 554 FVLNNSATDLNRHVEDAKGLLNLEKNANKOPLVTDSTATVGLRLKLGWVSTKNGT 613
 QY 699 DPGPNTKTLGELIKYKGGKTADLTKNIGVADSTNSLTVKLAKLISDL-----D 752
 Db 614 KRESNOVKADELFTGAGAAFTVTSKSENG-----KHTIVSYAETKADGLEKDG 665
 QY 753 AVNTKTLTA-SDKVTVDSGNNTAKLQNGDLTFESKONTGATPA----- 793
 Db 666 TIKLAVDNGNTDNLVGVNNGTAFTKGG--FETVKTGATDADRGRVTKADATANDAKK 722
 QY 794 -----TNSKTIGVGLKFTDN--NGIALDGTYYITKDKVGRADK 831
 Db 723 VATVADVATAINSAAFTVTEMLTTSIDEDNPTDNGKDALKAGDYLTFKAGKNLKYKR 782
 QY 832 G---SLDKSKPYLDK-----DKLVG-----EVEITNGIN-----AG 861
 Db 783 GKNITFDLAKNLEVKTAKSDDLITIGGNTPTGTTATPPVNTITSTADGLNFAKETADASG 842
 QY 862 GKAI--TGLSNTLTD-ATNATGHTVQLGIVDSTDKTRAASIGDVLNAGFNLKNNDAND 918
 Db 843 SKNVLYKGIATLTLPESAGAKSHVDL--NVDAKTKSNASIEDVLRAIGNIGNGNNDV 900
 QY 919 EVSTDTVPT----- 929
 Db 901 YATVDTVFTDSTGTTVTYVTOKADGKADVIGAKTSVIRKDHNGKLETGCKLDANN 960
 QY 930 ----- 929
 Db 961 GATVSEDDCKDPTGLVTAKTVIDAVNKSQWRVTEGATAEFGATAVNMGNAFTVSGTS 1020
 QY 930 ----NGNATTAKVYDGRKASVAVDVN--DGTTHLTLGADGNKNOIGVKTTLTKTDAR- 983
 Db 1021 VNFKNGNATTAATVSKDNGINAKYDVNVGDGLKI-----GDCKIYADTTTLTVGKV 1074
 QY 984 ----GPKALNFVSNGDDKALINAKDIADNLNLTAGEIR-----NTGTAFTAO--- 1029
 Db 1075 SVAPAGANSVN-----NNKRLVNAEGLALNLSWTAARADKYADESEBETDQEVKAGD 1128
 QY 1030 --TFQV--KRYKEN----- 1039
 Db 1129 KYTFKAGKMLKYKOSKEDTYSLODITLGLTSTLGGTANGRNDPTGVINKDGLITLAN 1188
 QY 1040 ----GDDNDADTTIVGKD-----AKTNOV 1060
 Db 1189 GAAAGTDAISNGNTISVTKGISAGNKEITNVKSAKTYKDOTONTADETODKEFHAUVKNA 1248
 QY 1061 NPLUKLKGKGLDIOITKDG-----TYTFEGINTOSGLKAGANTITLNNNG--LSLKNRAGNE 1113
 Db 1249 NEEEFGKNGATVSAKTODNNKHVTIDV--AEAKVGGLEKDTGKIKLKVANDGDN 1305
 QY 1114 QIOVGADGVKFAKVVNGVGA-----GIDGTTRITRDEIGFAGTNGS-----LDKSKPHL 1163
 Db 1306 LLLV--DAIKGASVAKGEINAVITTDATTAOGTANERKGVVYVKSNGATATEIDK----- 1358
 QY 1164 SKDGINAGKKTITNIOSEIADONSNDVATGKIYDLKTELENKISSYAKTAQNSLHEFSV 1223
 Db 1359 -----KVAATV--GVAKAINDAATFVKV-----ENDSATI-----DDSP 1392
 QY 1224 ADEQGNFTVSNYSYSDTSKSDVTF--AGEGITTAKVNGV--VRGIDOTKGLTTP 1279
 Db 1393 TPDGAG-----DAIKAGDITLTKAGKMLKYKROGKNTTFPLANDLSKSAFTVSD 1441
 QY 1280 KLVGNNGK-----GIVDSONGONT--ITGLSNTLANVTNDKGSVTEOGK 1326

Db 1442 KISLGTNGKNVNTSDPTKGLNFAKDKTGDNDANIHNLNGIASLITLIDLNSGAI--TVLGGN 1500
 QY 1327 IIKDEKTPRAASIVDVLASAGFNIQ-----NGEAVDFVSTYDVVNPADGNATTAKYT 1378
 Db 1501 GLTDNEKKRAASVADVLNAGMNVNGVKKPASANNQVNEIDFVATYDVDFSGKDTSTYAT 1560
 QY 1379 YD--DTSKTSKYVYDVVNDTTEVADKGLGVTTLLTSG-----TGAN-KFALSNOAT- 1430
 Db 1561 VESKDKNGKREV-----KIGATSVYIKDHNGKLTFTGKELKDNNGVTV 1604
 QY 1431 -----GDALYKASDIYVAHLNLSGDIOTAKAGASQAN-----SSAGYVADAGNKYV 1475
 Db 1605 TETDGEDEGGLVTAKAVIDAVNKKAGMRVKT--GANGQNDFFATVASCNTVFPADNGTT 1663
 QY 1476 YDSTDNKYYOAKNDG--TVDKTEVAK-----DKLVAQA--QTPDGLTLOANVKSVIN 1524
 Db 1664 AEVT-----KANDGSITVAKYVNVADGLKLDGDKIYADTTVLTVDGKVTAPN----- 1711
 QY 1525 KEQVNDANKKQINEDNAPVKGLEKASDOKTKNAAVTVGDULAAVAOTPLTFAGDGTGTTA 1584
 Db 1712 ----NGDGK-----FVDASGLDALNKLKLSWTA--TAGKESTGEVDPANSGO-----E 1754
 QY 1585 KRLGETLTIKGOTDINKLTDNNGIVAGTDFVTLAKDLNLSV-----NAG-----G 1635
 Db 1755 VKAGDKVTFKAG-----DLKIKQSGKDPYLSLKRKLKLTSEVFPDANGTGSSES 1805
 QY 1636 TKIDDKGVSEFVSSGQA-----KANTPVLSANGLDIGKAVISVNGVGTK-----DT 1681
 Db 1806 TKITKIDGLITTPANGAGAGANANTISVTKDISAGNKAVTNVVSGLKFKFGDHTLANG 1865
 QY 1682 DAANYOOL--NEVRNLGLGNAGNDN-----ADGNQVINAD-----IKKDPNNGS- 1724
 Db 1866 TVADFEKHIDNAKDLTNLDEKADNNPYADTATVGLDRGLGVNISDKTTGEPNOE 1925
 QY 1725 ----SSNRTVIKAGT-----VLGG-----KGNDEKLTATG-----V 1753
 Db 1926 YNAVNRNANNEVFKFSGGINVSGKTLNGFVITFELAKGEVKSNEFTVYNNADGSETNLY 1985
 QY 1754 QVC-----VDKGNANGDLSNWWVKTOK-----DSKKALLATYNNAGQTNLY 1796
 Db 1986 KVDMDYKSEDDIDPATSKPMTG--RTEKYKVENGVKVSANGKTEVTLTKNGSG--YV 2039
 QY 1797 ENNPAEADIRINOGIRFPHVNDGNEPYYOGRNGIDSSASGHSVAIGFOAK----- 1849
 Db 2040 TGN--OYADAIAKSGFEL-----GLADAABEAKARRESKDKQSLSKDRA 2081
 QY 1850 ----ADGEAAVALIGROTQAGNOSIAIGDNAQATGQ-----SIAIGTVNVAKGHS 1897
 Db 2082 ETVNANDKVFANGLMTKV--SAAFTVESTDANGDKVTTTFVKTDELPLTQIY---NTD 2135
 QY 1898 AIDDPSTVAKADNSYVGNNOPTDATOTDYFGYGNNTITYTESNSVALGSALSASGTHAG 1957
 Db 2136 ANGNKIVYKADGRW-----YELNADGTASKEVTLGVNANGKK----- 2174
 QY 1958 TOAKSDGTAGTGTTAAGATGVKFGAOTVAGAVSG-----ASGAERRIONV 2005
 Db 2175 -VYKVBENGADKRYTNADGAADKTGGEVSNDRVSTDEKHYVLLDRPNOSNGGVYIDV 2233
 QY 2006 AGEVSATSDAVANGSOLYKATGIGIAN--ATNELDHRHONENKANAGISSAMAMASP 2062
 Db 2234 ANGEISATSTDALINGSOLYVAKGVNINLAGOVNNLEKGVKVKRADAGTASALASQLP 2293
 QY 2063 QATIPGRSMYTGIAIHNGGAAVAVGLSKISDNGQVYFKINGSADPDQGHGAAGVAGFHR 2122
 Db 2294 QATMPKSMVALIAGSSYOGNGIAGVSRISDNGKVIIRLSGTTNSOGKTGVAAGVGYOW 2353

RESULT 5.
 US-08-913-942-4
 ; Sequence 4, Application US/08913942
 ; Patent No. 6200578
 ; GENERAL INFORMATION:

APPLICANT: St. Geme, Joseph
 APPLICANT: Barenkamp, Stephen J.
 TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Flehr Hochbach Test Albritton & Herbert LLP
 STREET: Four Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentln Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/913,942
 FILING DATE: 29-DEC-1997
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/409,995
 FILING DATE: 24-MAR-1995
 PRIOR APPLICATION NUMBER: PCT/US96/4031
 APPLICATION NUMBER: PCT/US96/4031
 FILING DATE: 22-MAR-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Vance, Dolly A.
 REGISTRATION NUMBER: 39,054
 REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELETYPE: 910 277299
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2353 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-913-942-4

Query Match 11.1%; Score 1185.5; DB 4; Length 2353;
 Best Local Similarity 22.7%; Pred. No. 8.3e-61;
 Matches 613; Conservative 330; Mismatches 832; Indels 925; Gaps 116;

QY 1 MNHLYKFNKATGTFFAVALAEYAKSHSTGGSCATGQVSVFRLSFARIALAVLIGAT 60
 1 MNKLFNVIMVMTQTVVVSSELTTRTKR-RASATVEFAYLATLLFNDQANAT----- 52
 QY 61 LNSAIVAGISGEADGGKGANARGDKSIAIGDIAQ-----ALGSOSIAIGDKNIHNSN 115
 DB 53 -----DEDEELDPVYRTAPVLSFHSDEGEGEKEVTENS 87
 QY 116 NNANIGAKASGNESIAIGDVLASGHSIAIGSDLLKKTVOISIELPIIRGQALN 175
 DB 88 WGIYFDNGK-----VLKAGAITLKAG-DNLKIKQNTDE----- 119
 QY 176 DIYADLTNLOKYRRTTAOGHASTAVGAMSYAKGHFSNACSTRATAGETYSLAVGLT-AT 234
 DB 120 -----STNASSF-----TSLKKNLDLDT 138
 QY 235 AKAASTAVGSNAQAIGFAATAVG-----GSTOVNLNREGIALGFGSOVLQKDNVYAN 288
 DB 139 SVAREKESFGANGDKRVITSDANGILAKTGNGVNHNL-----GLDSTLPDAVTNTGVL 193
 QY 289 VRAYAPDNDQIDNRKATPFNGATDVESIG-NSNGNDSIRKLIINAGASADDDAVNVA 347
 DB 194 SSSSTPND---VEKTRATYK-----DVLNAGWNKIG-----AKTAGSNV- 230
 QY 348 QLKRAVRL--ANROITFGKDSNN-----RVEKGLGKTLTTTGAQTSALTDHNGVQNG 401

DB 231 ---ESVDVSAVNNVEFIINGDKNTLIDVLLAKENGKTEVEKFPKTS-----VIREK 279
 QY 402 DGLKVLAEITLTLKAVTENTLANEKVYVGKRLITDIXIGFTNDMNGIDESKPYLDKOT 461
 DB 280 DG-----KLTGKENNDINKVT---SNFATDN---TDEONG-LVTAKAVID- 318
 QY 462 GHAGGOKITKLKAGVDDAATVYGOLKKVNOFAEALDTFYKVKDKGN-----DA- 514
 DB 319 AVNKAQRVATTTANGONGDFATVA--SGTNVFEESGDGTTSVKYKDNNGCITVYKADK 376
 QY 515 -----NDSKITIVGKNNKPDGTQVNTL-----KLKGENGVDTT 548
 DB 377 YGDGLKFDSDKKIVADPTLALTVAIGKVAELAKEDKKLVNMGDLVTALGNLSWKAKA 436
 QY 549 ETGTYTEGILNONGNGLTVGNSITLNDGLSVK-----NTNSKQIQVADGDTITFDISNKP 604
 DB 437 DTGDALE-GISKQEVKAGETVTFKAGKMLKVKQDANFTYSLODALTGLTSTIGGTN 495
 QY 605 GAGIENTRTITRGIGF--ANNISGLDANKPRLTPIGINAGGKELTNVOSATINPATNGQ 662
 DB 496 G-GNDAKIVYNNKGLITTPAGNGCTGTNTISYTKGICAKKAITNVASGLR-AYDAN 553
 QY 663 LDPMNRLST-----ANTE--KSGSAATIKDLYNLQVPLTFAG- 698
 DB 554 FDLNNSATDLNRHVEDAKKGLNLNEKNANKQPLVTDSTAAVGDRLKLGWVSTKNGT 613
 QY 699 DTGPNVTKKIGELLKVKGGKTTADLTKNNGVADSTNSLTVKLAKTSLD-----D 752
 DB 614 KESNOVKQADEVLFTGAGAAVTSKSENG-----KHTIVSVAEKTKADGKEKED 665
 QY 753 AVNTKTLTA-SDKVTVDSGNNTAKLQNGDLTFESKONTGTPA----- 793
 DB 666 TIKLVDNONTDNLVLTGNGNGTAVTKG--PEYVGTGATDADRGKVYKDAFANDADK 722
 QY 794 -----TNSKITGVNGLKFTDN--GIALDGTYYTKRVGFAKOD 831
 DB 723 VATVKDVAATINSATFVKENLTISIDEDNPNDKDKDALAKAGDILFKAGKNNLKVRKD 782
 QY 832 G---SLDKSKPYLDK-----DKLVG-----EVELTNGIN-----AG 861
 DB 783 GKNTFDLAKNLEKVKAKVSDTLTIGNTPGTGTPAPRVNITSTADGINKFAKETADAG 842
 QY 862 GKAI--TGLSNTLTD-ATNATGHTVQQLGYVSDTKTRAASIGDVLNAGFNKKNNGDAD 918
 DB 843 SKNVYIKGIAFTTLTFESAGAKSHVDL--NVDAFKSNAASIEDVLRAGMNIOGNGNVD 900
 QY 919 FVSTYDTPFI----- 929
 DB 901 YVATYDFVNTDSTGTTYVTYQKADGKADYKIGAKTSVKIDHNGKLFQKDLKDANN 960
 QY 930 +----- 929
 DB 961 GATVSEDDGKDTGTLVTKATVIDAVNKSQWRVTEGATAEATGAVNAGNAETVTSCTS 1020
 QY 930 -----NSNATTAKTATYDGKASKVAVDVNV-DGTTIHLTGADKNNNOIGVATTLTETDAK- 983
 DB 1021 VNRKNGNATTATYSKONGNINAKTIDVNVGDLKT-----GDDKILVADPTTLTGTGKV 1074
 QY 984 ---GDKAIFSVNSGDDKALINAKDIADNLNTLAGEIR-----NTKGTADTALQ--- 1029
 DB 1075 SVPAGANSVN-----NNKKLVABEGLATALANLSTAAADYKADDESSEGEIDQEVKAD 1128
 QY 1030 ---TFQV---KKVEN----- 1039
 DB 1129 KVTFFKAGKNNIKVQSKSEKDFYSLQDTLTLGLTSLTIGGTANGRNDGTVINKDGTLTILAN 1188
 QY 1040 -----GDDNDADRTITVGKD-----AKTNGV 1060
 DB 1189 GAAAGTIDASNGNTISYTKGISAANKETINVKSAKLTYYDTONTADETDOKFEHAAVKA 1248
 QY 1061 NTLKLGKNGLDIOTNRKD---TVTFPIGINTQSGLKAGNNTLNNNG---LSIKPTAGNE 1113
 DB 1249 NEVEFGKNGATVSAKTDNNGKHTVITDV---AEAKVVGLEKDDGKIKLKVNDTNGNN 1305

QY 1114 QIOGVAGVYKFAKVNNGVGA-----GIDGTRIRIDEIGFATNGS-----LDKSKPIL 1163
 Db 1306 LLEV--DARKGAVYKAFENAVTTDAFTTAQGTANERKGVYVKGSGATFETDK----- 1358
 QY 1164 SKDINAGGKKTINIOSGEIAONSNDVAVGKIYDLKTELENKISSTAKTAONSLHEFSV 1223
 Db 1359 -----KKVATV--GDVAKAINDAATFVKV-----ENDDSATI-----DSDP 1392
 QY 1224 ADEQGNFTVSNPSSYDTSKISDVITF--AGENGITTKVKNKV--VVRGIDOTKGLTTP 1279
 Db 1393 TDDGAN-----DALKAGDILTLLKAGKMLKVRKRGKNTFALANDLSVKSATVSD 1441
 QY 1280 KLFVGNNGK-----GIYDSONGNT-----ITGSNTLANTNDKSGVRTEQOK 1326
 Db 1442 KLSLGTNGKAVNTTSDTKLNFPAKDSKTGDDANIHNLNGIASLTLPDLNLSGAT--TLGNG 1500
 QY 1327 IIKDEKTRAASIVDVLASGFNLOG-----NGEAVDEVSTYTVNPAAGNATTAKYV 1378
 Db 1501 GITDNEKKRAASVYKDVNLNAGMNYRGVYKPA SANNOVENIDFVATYDTVDFVSGDKDTTSV 1560
 QY 1379 YD--DTSKTSKVYDVVNDVDTTEVDDKLGVTTLTSTG-----TGAN--KRALSOAT- 1430
 Db 1561 VESKDNKREY-----KIGAKTSYIKDHNGKLTFTGKELKDKANNNGVTV 1604
 QY 1431 -----GDALYKASDIVAHLNLSGDIOTAKASOAN-----SSAGYVADADGNKYI 1475
 Db 1605 TETDGDGEGNGLYTAKAVIDAVKAKGRYKTT--GANGQNDPRTAVASGNVTFADNGT 1663
 QY 1476 YDSTDNKYYOAKNDG--TYDKTEVAK-----DKLVAQA--QTPDGLAOMNYSVIN 1524
 Db 1664 AEYV-----KANDGSITVYKYNKVDGLKLDGDKIVADTVTLVADGKYTAPN----- 1711
 QY 1525 KEQVNDANKKOGINEDNAPYKGLEKASDNKTKNAVTVYGDNLNAVQOTPLTFRAGDTGTA 1584
 Db 1712 -----NDGKK-----FVDASGLADALNKLISWTA--TAGKEGTGEVDPA NSAGQ-----E 1754
 QY 1585 KRIGETLITIGQOTDNKLNDNIGVAGDGTGFKLADLTFLNLSV-----NAG-----G 1635
 Db 1755 VKAGDKVTFKAG-----DNLKIKOSKDPSTSLKKEKLDLTVEKERDANGGTSFS 1805
 QY 1636 TKIDDKGVSPDSSQA-----KAMPVLSANGLDGKGVISNVGKGTK-----DT 1681
 Db 1806 TKITKDGILTPPANGAACANTANTISVTKDISAGNKAVTNVYSGLKKEFGDHTLANG 1865
 QY 1682 DANNVOL--NEVRNLLGNGNNDN-----ADGNOVNIAD-----IKKDPSSGS----- 1724
 Db 1866 TVADFEKHYNAYKDLNDEKAGDNNPTVADNTAATVADLRLGLNVAISADKTTGEPNOE 1925
 QY 1725 -----SNRTYIKAGT-----VLG-----KGNNDTEKLTATG-----V 1753
 Db 1926 YNAQVNRANANVYKFSNGNIVSGKTLNGTRVITFELAKEGVYKSNFETLYKNADGSTNLV 1985
 QY 1754 QVG-----VDKGNANGDLSNVVKTOK-----DQSKKALATYNAAGOTNYL 1796
 Db 1986 KYVDMYSKEDIDPATSKPMTG--KTEKYKENGKRVYVANGSKTEVTLTNKSGS--YV 2039
 QY 1797 TNPAEALIDINOGIRFHVNDGNOEPVYVQNGNIDSSAGSHSAIGFQAK----- 1849
 Db 2040 TGN--OVAADIAKSGFEL-----GLADAAEAKAPAESAKDKOULSKDKA 2081
 QY 1850 -----ADGEAVALIGROTOAGNOSIAIGNAQATGDO-----STAIGUNVAVAKHSG 1897
 Db 2082 EYNAHADKVFANGLNTKV--SAATVESTDANGDKVTTTVEYTDVELPLTOYV--NND 2135
 QY 1898 AIGDPSTVKAANSYVSGNNQTFDQTQDVFGVGNNTITVESNVALGNSAISACTHAG 1957
 Db 2136 ANGNKIYKKRADGKW-----YELNADGTASNKVEYTLGVNDANGK----- 2174
 QY 1958 TOAKKSOGTAGITTTAGATGVAGGQTAVGAVSIG-----ASGAERIQNV 2005
 Db 2175 -VYKVTENGADKWYTYTNADADAKTKGEVSNDSKYSTDEKHVYRLDPNNSNGKGVVINDV 2233

QY 2006 AAGEVATSTDAVNGSQYKATOGIAN---ATNELDRHIONENKANAGISSAMAMAMP 2062
 Db 2234 ANGELSATSTPAINSQYIXNAKGVNTLNGOVNLEGVNKGRRADAGTASALASQPL 2293
 QY 2063 QAVTGRSMVTCGIATNGGCAVAVAGLSKSDNCQWFKINGSDTQGHVGAAGACFHP 2122
 Db 2294 QATMGKSMVALIAGSSYOGNGLAIGVSRISDNGKVIIRLSGTNSOGKGTGVAAGVGYOW 2353
 RESULT 6
 US-09-669-974-33
 : Sequence 33, Application us/09669974
 : Patent No. 6333173
 : GENERAL INFORMATION:
 : APPLICANT: PEAK, Ian Richard Anselm
 : APPLICANT: JENNINGS, Michael Paul
 : APPLICANT: MOXON, E. Richard
 : TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
 : FILE REFERENCE: 065064/0128
 : CURRENT APPLICATION NUMBER: us/09/669,974
 : CURRENT FILING DATE: 2000-09-26
 : PRIOR APPLICATION NUMBER: US 09/377,155
 : PRIOR FILING DATE: 1999-08-19
 : PRIOR APPLICATION NUMBER: PCT/AU98/01031
 : PRIOR FILING DATE: 1998-12-14
 : PRIOR APPLICATION NUMBER: GB 9726398.2
 : PRIOR FILING DATE: 1997-12-12
 : NUMBER OF SEQ ID NOS: 33
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 33
 : LENGTH: 2353
 : TYPE: PRT
 : ORGANISM: Haemophilus influenzae
 US-09-669-974-33

Query Match 11.1%; Score 1185.5; DB 4; Length 2353;
 Best Local Similarity 22.7%; Pred. No. 8.3e-61;
 Matches 613; Conservative 330; Mismatches 832; Indels 925; Gaps 116;
 QY 1 MNHITYKVFNKATGTFMAVAEYAKSHSTGGSCATGOVGSRTTSPARIALAVLVGAT 60
 Db 1 MNKIFNVINWNTQVWVYVSELTTRHTK--RASATVETAVLATILFATVQANAT----- 52
 QY 61 LNSAVAGIGISEADCGKGANAGDKSIAIGDIAO-----ALGSQSIAGDNIIVNSN 115
 Db 53 -----DEDELDPPVYKRAPVLSFSDSGEGTEKEVTENSN 87
 QY 116 NNANIGARASGNESIAIGDVLASGHSIAIGSDLYLKETEVOQISELPIIRGOKALN 175
 Db 88 WGIYEDNKG-----VLKAGATLLKAG--DNLKIKONTDE----- 119
 QY 176 DIYQIADINLOKRYRTAOGHASTRAVAGMSYAKGHFSNAPOTRATAGTSLAVGLT-AT 234
 Db 120 -----STNASSF-----TSLKRDLDLT 138
 QY 235 AKAAASSIVAGSNAOIGPATAVG-----GSTOVNLRGIALGFGSOVLQKNDVNAAN 288
 Db 139 SVATEKLISFGANGDKVDITDPAINGLKLAKTGNVHNLN-----GLDSTLPRAVNTGVLS 193
 QY 289 VRAVAPDDNOPIDNRYKATFFKNGATDVSIG--NSNGNDSIRKTIINVGASADTDAAVNA 347
 Db 194 SSSFTPPND--VEKTRATVYK-----DVLAQMNKIG-----AKTAGSNV- 230
 QY 348 QLKEAVRL--ANKQITFFGDDSN-----RVKGLGKLTITTGGAQTSALDHNIGVQNG 401
 Db 231 ---ESVDLSAVANNVEFTTGDKNTLDVLTAKENGKTEVEFTPTS-----VIREK 279
 QY 402 DGLKQVLAETLSLKMVTENLTANEKVYGTGRTLTDTKIGFTDMNGIDESKRYLKDPT 461
 Db 280 DG-----KLTFGKENNDTNKVT--SNATPDN--TDEGMLYAKAVID-- 318
 QY 462 GIHAGGOKITKLTAGVVDADAATYQOLKKVNOQTAEASALQTFYVKKVDKNGN-----DA- 514

Db 319 AVNAGRWRTTANQONCFATVA--SGTNVFESEGDGTATVTKDFNNGITVVKYDAK 376
 QY 515 -----NSKTIITVCKNNKPDGTOYNTL-----KLKGEQDVYTT-----548
 Db 377 VGDLKEDPSKRIYADTTALTALVTGGKVAELAKEDDKKRLVAGDLVYALGNLSMKAKAEA 436
 QY 549 ETNGTVEFGUNONGITVGNSTLNNDLNVK-----NTNSKQIOVAGDITFEDISNSKP 604
 Db 437 DTGDALE-GISKODEVAGETVTFKAGKNLKVAGDGNFTYSLDALTGLTSTLIGSTTN 495
 QY 605 GAGIENTRITRDIGF--ANNTGSLDANKPRLPTGINSAGKELTVNOSAINPATNGQ 662
 Db 496 G-GNDATVINKDGLITTPAGNGTGTNTISTVKDQIKRGNKALITVVASGLR-AVDAN 553
 QY 663 LDFPNRST-----ANTE-----KSSNATIKDLINLSOYPLTFAG- 658
 Db 554 FDLVNSATDLNRHVEDAYKGLNLNENKANKOPLYDTSTAAATVGDRLKLGWVYSTKNGT 613
 QY 699 DTGNVTKKIGELIKVKGKTTADLTFRKNNIGVADSTDSLTVYKLAKTLSL-----D 752
 Db 614 KEESNOVKODEVLTFTAGALTYTSKENG-----KHTTVSAETKADCGLEKDG 665
 QY 753 AVNTKTTLTA--SDKVTYVDSGNNTAKLQNGDLTFKONTGATPA-----793
 Db 666 TIKLVKNOMTDNVLTVGNNGTAVTKG---FETVKGTADNRGKVTYKDATANDADK 722
 QY 794 -----TNSKTIGVDLKFIDN---NGIALDGTYYTRKQVFAKOD 831
 Db 723 VATVKDVAATINSAAETVKTENLTTSIDENPTEGKODALKAGDITLTFKAGKMLKVRD 782
 QY 832 G-----SLKSKPYLDK-----DKLVK-----EVELTNGIN-----AG 861
 Db 783 GKNTFPLAKLEVKAKVSDTLTIGNTPTGTTAPKVNITSTADGLNFAKETADAG 842
 QY 862 GKAI--TGLSTLTD--ATNATTVHVLQGIYDSTDKTAAISIGVLAGENLKNNGDADK 918
 Db 843 SKNYLKGIAITTLTEPAGAKSSHVDL--NVDATKKSMAISIEVYLAAGNIGNGNVD 900
 QY 919 FVSTYDVTDF-----929
 Db 901 YVATYDVNFTDSTGTTVTVTKQADKGADYKIGAKTSTYKDHNGKLTGKDLKDANN 960
 QY 930 -----929
 Db 961 GATVSEDDGKDTGLVTAKTVIDAVNKSGRVTGEGATETGATAVNAAGNAETVTSCTS 1020
 QY 930 -----NGNATTAKYVDGKASKAVAYDVNV-DGTTIHLTGADGNKQOIGVTTTLTKTDK- 983
 Db 1021 VNFKNAGNATTAATVSKDGNINVKYDVAVGDSLKI-----GDOKKIYADTTTLTVTGKV 1074
 QY 984 -----GDKAINEVNSGDDKALINKADIADNLNTLAGEIR-----NTKGTADTALO--- 1029
 Db 1075 SVPAAGANSVN-----NKKKLVAEGELATALNLSMTAKADKYADGESEGETDEVKAGD 1128
 QY 1030 -----TEOV---KVKEN-----1039
 Db 1129 KVTFRAGKNLVKQSEKDFYSLDPTLGLTSLTGSTANGRNDGTIVINKDGLITPLAN 1188
 QY 1040 -----GDDNDADATITVGDK-----AKTNOV 1060
 Db 1189 GAAGTASNGNTISTVTKDISAGNKEITVYKSAKTYKQDONTADETQDEPFAAANKA 1248
 QY 1061 NTLKLGKNGIDTQNKDG---TVTFGINOSGLKAGNNPTLNNNG---LSIKTAGNE 1113
 Db 1249 NEVEFVNGATVSAKTDNNKHTVYIDV--AAKVGDDGLEKTDKIKLKDNTGNN 1305
 QY 1114 OIOVAGDVKRAKYNNGVYGA-----GIGGTRITRDEIGFAGTNGS-----LDKSPHL 1163
 Db 1306 LITV--DATKASVAKGEFNAVTTDATTAQSTNANENKRGKYVVKSGNATLATETDK----- 1358
 QY 1164 SKDINAGKRTINOSGELIAONSNDVATGKIYDLTKTELENKISTSTAKTAONSIEFSV 1223

Db 1359 -----KKVATV--GDVAKAINDAATPVKV-----ENDDSATI-----DDSP 1392
 QY 1224 ADEQNNFTVSNPSSYDTSKSDVTF--AGENGITTKVKNV---VRYGIDQTFGLTTP 1279
 Db 1353 TDDGAN-----DALKAGDTTLTKAGKNLKVRRGKNITRALANDLSVKSAVSD 1441
 QY 1280 KLVGNMNNK-----GIVDSONGQNT---ITGSLNLANVYNDKGSVRTTEQK 1326
 Db 1442 KLSLGTNGKNVNTISPTKCLNFAKDSKTCDDANHINGLASTLTLTLNSGAF--INLGN 1500
 QY 1327 ITKEDKTRASIVDLSAGFNLC-----NGEAVDSTYDTVFADGNATTAKYT 1378
 Db 1501 GLTDEKKAASVQVNLNAGWVNGVGPASANNQVENIDFVATYDTVFVSGDKDTSTYT 1560
 QY 1379 YD--DTSKTSKYVYDVNVDDTTEVKKLGVKTTTLTSTG---TGAN-KFALSNOAT- 1430
 Db 1561 VESKDNKRPTEV-----KITAKTSYIKDHNGKLTFLKELKANNNGVTV 1604
 QY 1431 -----GDALVKASDIVAHLNLTSLGDIOTAKGASQAN-----SSAGYVADGNKY 1475
 Db 1605 TETDGDENGGLVTAKAVIDAVNAKAGWRTT--GANGQNDDEATVASTNTVFAAGNGTT 1663
 QY 1476 YDSTDKKYYQAKNDG--TYDKTEYAK-----DKLVQA--QPPDGTLAOMNYSVIN 1524
 Db 1664 AAVT-----KANDGSITVKNYKVDGLKLDGDKIVADTVLTVAAQKVTAPN----- 1711
 QY 1525 KEQVNDANKKQGINEDNAVYKGLERASDNKTKNAAVYGDILNAAQPTLTAGDTGTA 1584
 Db 1712 ---NDGKR-----EVDASGLADLNKLSTMTA--TGAKGTEGVDNPSAGO---E 1754
 QY 1585 KLGELTTRKGGOTDINKLDNNIGVAGTGDGFTVLAIDLNLNSV-----NAG---G 1635
 Db 1755 VKAGDVYTKRAG-----DNLIKQSGKPTTSLKKEILDLSVERKDNAGGTGSR 1805
 QY 1636 TKIDKGVSEVDSGQA-----KANTPVLSANGLDGKRVISNKGTK-----DT 1681
 Db 1806 TKITKGLDITTPANGAGAGANTANTISTVKQISAGNKAIVNVSGLTKPFQDGLTIANG 1865
 QY 1682 DANNVOOL--NEVRNLIGLGNAGNDN--ADGNQVNIAD-----IKKDNSSG--- 1724
 Db 1866 TVADFEKHVDNAKDLTNIDEKADNPTVADNTAATVADLGLGVLNADTTGEPNOE 1925
 QY 1725 -----SSNRTVTKAGT-----VLG-----KGNNDTEKLATG---V 1753
 Db 1926 YNAQVRNANNEVFKSSGNGINWSKTLNGRVITFELAKEBYKSNFTYKANADGSTNLY 1985
 QY 1754 QVG-----VDKGNANGDLSNVYKTK-----DSSKALLATYNAAGQNTYL 1796
 Db 1986 KVGDMYTSKEDIDPATSKPMTG---KTERKYKENGKVSANGSKTEVTLTKNGSG---YV 2039
 QY 1797 FNPFAIDRINEQIRFHVNDGNQEPVVOGRNGIDSSASGKHSVAIGFOAK----- 1849
 Db 2040 TGN--GVADIAKSGFEL-----GLADAEEKAFESAKKQOLSKDKA 2081
 QY 1850 -----ADGEAAVAIGROTQAGNOSIALGDNAOATGDO-----STAIQTGVNVAGKHG 1897
 Db 2082 ETYNAHDKYRFAAGLNTKY--SAATVESTDANGDKVTTTPKTYDVELLTQY---NTD 2135
 QY 1898 AIGDPSTVRKADNSYSVGNNOFTDATQTDVFGVNNITVYESNSVALGNSNAISAGTIAG 1957
 Db 2136 ANGNKTIYKADGKW-----YELNADGTASKEVTLGIVNDANGK----- 2174
 QY 1958 TOAKKSDGTGTTTATGATGTVKGFAGQTAAGVAVSG-----ASGAERIONV 2005
 Db 2175 -VVKVTENGADKYYTNAGDADKTGGEVSNKVSSTDEKHVVRLDPNNOSNKGVIYDVA 2233
 QY 2006 AGEVASTDAVNGSOLKATQIGIAN--ATNELDRITHONENKANAGISSAMAMASRP 2062
 Db 2234 ANGEISTSTIDALINGSQLTAVAKGYTNLAGOVNLEBKYNKKGKRRADGATATALAASQLP 2293
 QY 2063 QAVIPGRSVYTGIAITHNGOGAAVAVLSKLSUNGQWVERKINGSADTQGHVAAVAGAFHE 2122
 Db 2294 QATMPKSMVAIAGSSYOGNGIAIGVSRISDNGKVIIRLSGTTNSGKGTGAAGVGYOW 2353

RESULT 7

US-09-268-347-47

Sequence 47, Application US/09268347

Patent No. 6335182

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS

FILE REFERENCE: 1038-860

CURRENT APPLICATION NUMBER: US/09/268,347

CURRENT FILING DATE: 1999-03-16

NUMBER OF SEQ ID NOS: 54

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 47

LENGTH: 2354

TYPE: PRT

ORGANISM: Haemophilus influenzae

US-09-268-347-47

Query Match 11.0%; Score 1177; DB 4; Length 2354;
 Best Local Similarity 23.0%; Pred. No. 2.6e-60;
 Matches 612; Conservative 327; Mismatches 876; Indels 846; Gaps 114;

QY 1 MNHKKVFNKATGTEPMVAEYAKSHSTGGSCATGQVSVTLTSPARIALAVLIGAT 60
 1 MKKIFVFNWMTQTMVYVSELRTHTK--RASATVETAVLATLTFATVQANAT----- 52
 DB 61 LMSAYAGIGISEADGGCGANARBDKSLAIGDIAO-----ALGSQSLAIGDNKIVHNSN 115
 DB 53 -----DEDELDPPVVRAPVLSFHSDKSGTGEYKTEVNSN 87
 QY 116 NNANIGAKAGNESTALIGDVLASGHASIALGDDLYLKKETVQOISELLPIIRGOKALN 175
 DB 88 WGIYEDNKG-----VLKAGATILKAG-DNLIKIKONTDE----- 119
 QY 176 DLYOLADTLMQYRRTHAOGHASTAVGANSYAKGHFSNAFGTRATAREGYSIANGLT-AT 234
 DB 120 -----STNASSP-----TYSLKQDLTDLT 138
 QY 235 AKAASTIAVGSNAQAIGPATAVG-----GSTOVNLNGIALGSGOYLKQNDVNAAN 288
 DB 139 SVATEKLSFGANGDKVKITSDANGKLAKTNGNVLN-----GLDSTLPDAVTMTGYLS 193
 QY 289 VRAVAPDDNQPIDNRKATFKGATDVSIG-NSNGNSIRKKIINVAGSADTDAVVNA 347
 DB 194 SSSFTPNPND--VEKTRATIVK-----DVLNAGNNIKG-----AKTAGNV- 230
 QY 348 OLKEAVRL--ANROITFEKGDSDSN-----RVEKGLKTLITITGAQTSALTTHNIGVONG 401
 DB 231 -----ESVDLSAVANNVEFITGDKNTLDVLTAKENGKTEFEVKTPPTS-----VIREK 279
 QY 402 DGLKVALAETLSLKKVVTENITANEKVYVGTGRLTTRDKIGTNDMNGIDESKPLDKDT 461
 DB 280 DG-----KLFTEKENNDNRKVT--SNRATDN--TDEGNGLYTAKAVID-- 318
 QY 462 GIHAGGOKITKLAVGVDDADATYGLKKVNOTAASALQTFVKKVKKGN-----DA- 514
 DB 319 AVNKAAGRWKTTTANGQNDPAFIVA--SGTNVTFESGDTTASVTKDINGNGLTYKYIDAK 376
 QY 515 -----NDSKLITVGNKNNKPDQOVNTL-----KLKNGENVDTT----- 548
 DB 377 VGGGLKFDSDKKIVADTTLTLYTGKVAELAKEDDKKILVNGDLYTALGNLSWKAKAEA 436
 QY 549 ETNGVTFEGLNDNGLTVGNSLNDGLSVK-----NTNSNKQIOYGAGDITPTDISNSKP 604
 DB 437 DHDGALF--GISKDOEAKGEVTFKAGKNLKYODGANTFYSLODALGTLITLGGTIN 495
 QY 605 GAGIENTTITTRDIGF--ANNUGSLDANKPRLPTGGINAGKELTNVQSALNPTNGQ 662
 DB 496 G-GNDKATVINDGLTITPANGGTTGTNTSVTKDGIKAGKAKAITNVAAGLR-AVDAN 553

QY 663 LDFMNRIST-----ANTE-----KSGSAATIKDLINISOVPLTFAG- 698
 DB 554 EDVLLNSATDLNRHVEDAYKGLINLNKNNANKQPLVDTSTAATVGLRKGWVVSSTKNGT 613
 QY 699 DTGBNVTKKIGELIKVKGCTTADDLTKNNIGVAASTDSLSVLAKLTSLD-----D 752
 DB 614 KEESNOVKQADEVLEFGAGATVTSKSENG-----KHTITVSAEIKADCGLEKQD 665
 QY 753 AVNTKTLTA--SDKVTVDGNNNTAKLQNGDLTFESKONTATPA----- 793
 DB 666 TIRKAVNONDNLVLYGNGNTAVTKG--FEVTKGATDADRGVYKATANDADAK 722
 QY 794 -----TNSKITGVGLKFTDN--NGIALDGTYYTKKQVGAQOD 831
 DB 723 VAVKDVATATINSATFVKTEENLTTSIDEDNPTDNGKDALAGDILTFKAGNKLKVRKD 782
 QY 832 G--SLDKSRPYLDK-----DKLKVG-----EVELTINGIN-----AG 861
 DB 783 GKNTTFDLAKNLEVKTAQVSDTLTIGGNTPTGTTATPKVNIITSIDGLNFAKETADASG 842
 QY 862 GKAI--TGLSNTLTD--ATNATTHVTLGIVDSTDKTRASIGDVNAGENLNKNGDAND 918
 DB 843 SKNYLKGIAITTLTLEPAGAKSSHVDL--NVDAATKKSNAASIEDVLRAGNNGGNNND 900
 QY 919 FVSTYDTVDPL----- 929
 DB 901 YVATYDTVNEFDSTGTTVTVQKADGKADYKIGAKTSYIKDHNGKLTFGKLDLANN 960
 QY 930 ----- 929
 DB 961 GATVSEDDGKDTGGLTAKTVDAVAVKSGWRVTEGATFETGATAVMAGNAETVSGTS 1020
 QY 930 -----NGNATTAQVYDGRKASRVADVNV-DGTTIHLTGADGNKNOIGVKTTLTKTDK- 983
 DB 1021 VNFKNAGNATTAQVYDGRKASRVADVNV-DGTTIHLTGADGNKNOIGVKTTLTKTDK- 1074
 QY 984 -----GKRAINESVSGCDKALLAKADINLNTLAGETR-----NTKGTATALO-- 1029
 DB 1075 SVFAGANSVN-----NNKILVNAEGLATLNLNLSWTAKADYKADGESBETQEVKAGD 1128
 QY 1030 --TFQV--KVKEN----- 1039
 DB 1129 KVTFRAGNKLAVKQSEKDFYTSIADDTLGLTSLTGLGTANGRNDGTIVIKKDLITLAN 1188
 QY 1040 -----GDDNDADTLTGVK- 1060
 DB 1189 GAAGTASNSNTISVTKDGSAGNKEITNVKSAKTKYKDTONTADETOKEFHAAYKNA 1248
 QY 1061 NTLKLGKNGIDTQTKNDG--TVTFGINTOSGLKAGNNTLNNNG--LSIKTRAGNE 1113
 DB 1249 NEVEFVGKNGATVSAKTDNNNGKHTVTDV--AAKAGDGLKEDTDKIKLKVNDTDGN 1305
 QY 1114 QIOVGADGVKFAKNNNGVGA-----GIDGTYRITDEIGFAGTNGS-----LDKSPHL 1163
 DB 1306 LFTV--DHTKASVAKGEFNAVTTDATTAOSTNANEGKAYVAGSKNATATEIDK----- 1358
 QY 1164 SKDGINAGKKTITNIOGSELAONSNDNAVTEGKIYDLTELENKISSTAKTAQNSLHEFSV 1223
 DB 1359 -----KKAATV--GDVAKAINDAATFVKV-----ENDSARTI-----DDSP 1392
 QY 1224 ADEGNNFTVSNPYSSTDSKTSVITF--AGEKQITTKVKGV--VRVGDITQKGLTTP 1279
 DB 1393 TDDGAN-----DALAGDTLLKLGKKNLKVRRDKNITFALANDLSVSATVSD 1441
 QY 1280 KLVGNNGN-----KGIYI--DSONGONT-----ITGLSNTILANVTNDRGSVTEBOG 1326
 DB 1442 KLSGTNNKNVNTSDTGLKFAKDSKTGDANIHNLGINSITLTLNSGAT--TNLGCN 1300
 QY 1327 IIKDEKTRASIVDLVSAGFNLOG-----NGEAVDFVSTYDVYFNADGNAATPAKY 1378
 DB 1561 GITNEKRRASVADVLNAGNNGVGPAPASANNQVEHIDVATYDTVDVSGKODTTSVT 1560
 QY 1379 YD--DTSKTSKVYVYVNVVDTTIEVKDKLGVKTTTLTSTG--TGAN--KFAISNOAT- 1430

```

Db 1561 VESDNGKRTV-----KIGAKTSVJIKDHNGKLPFGKELKLDANNNGVTY 1604
Qy 1431 -----GDALVYASDIVAHLNLTSGDIQPAKASQAN-----SSAGVADAGNKVI 1475
Db 1605 TETDQKDEGNGLYAKAVIDAIVNKAQMRVKT--GANGQNDQFATVAGTIVTFADNGTT 1663
Qy 1476 YDSIDNKTYQAKNG--YDQKTEYAK-----DKLYAQ--QTPQSTLAQMMVKSVIN 1524
Db 1664 AEVT-----KANDGSLTVKYNVADQKLKDGKIVADTVTLVAVAGKYAPN----- 1711
Qy 1525 KEQVNDANKKQGINEDNAFVGLKKAASDNKTNAATVYGLNVAQPTLFEADDTGTTA 1584
Db 1712 -----NGDGR-----FVDSGLADLANKLSMTA--TAGEGTEGEVDPAASQ-----E 1754
Qy 1585 KRLGELLTKGQDTNKLNTNNICVAVGTGTFVKLAKDLTLNLSV-----NAG-----G 1635
Db 1755 VKAGDKVTFKAG-----DNLIKQSGKDFYLSLKLKDLTSVEFKANGGTSSES 1805
Qy 1636 TKIDDKGVSEVDSGQA-----KANTPYLSANGLDLGKVTLSNKGKTK-----DT 1681
Db 1806 TKITKDGTLTPANGAGAGANTATISVTKDQISAGNKATVNVVSGLKKEFGDHTLANG 1865
Qy 1682 DAANQOL--HEVRNLGLGNAGNDN-----ADGNQVNIAD-----IKKDPNNGSSS-- 1726
Db 1866 TVADEKHTDNYKDLTLNLDKDGADNPTVADNTPATVGLRGLGWTISADKTTGERPQE 1925
Qy 1727 -NRVYKAGTV--LGKGNNDTEKLTATGVOGVYDKNAGNDLSNVVTKQKSGKAL 1783
Db 1926 YNAQVRNANNEVKFGSNGINVSQKLDNGTGVITPELAKGEVKSNEFTVKNADQSEITNL 1985
Qy 1784 LAT--YNAAGTNLTNNPAA--IDRINEQIRFFHYNDGNEPV-----YQGRNG 1831
Db 1986 VKVGMYSKEDIDPATSKEPTGTEKYKENGKVSANGSKTEVTLTKKSGVYTGNOV 2045
Qy 1832 IDSSAGCHSVAIGQAKADGEAAVAIGRQTAQGNOSIAIDNMAQATDQSIAGTGMVY 1891
Db 2046 ADAIA--KSGPELGADAAEKAFAESAKQKLSKDA--ETVAHAKVAFANLNTKV 2101
Qy 1892 AG--KHSGAIGDPST--VKADNSYSVGNNOFTDATQTFVGVNITVT-----ES 1939
Db 2102 SAATVESTDANGDKVTTFVKTDELPL-----TOIYNTPD--ANGNKIVKADQWEL 2153
Qy 1940 NSVALGNSAISAG--THAGTQAKKSDGTAGTTTACATGIVGFAQOTAVGANSV-- 1994
Db 2154 NADGTASKEVTLGVNDANGKRVKVTENGADKMYTINADGAADKTKGEVSDKVTDEK 2213
Qy 1995 -----ASGAEIRIQNVAGVSAVSATSDAVNGSOLYKATOGIAN--ATNELDRI 2041
Db 2214 HVRVLDPNNOSSNGKGVLDNNAANGELISATSTDAINGSALYAVAKVYTLNLAGOVNLEBKV 2273
Qy 2042 HONENKANAGISSAMAMASMPQATYIPGRSMYTGATINHGOGAVAVGLSKLSDNGQWEK 2101
Db 2274 NKVGRADAGTASALASQLPQATMPKGSMAVAGSSVQOGMGLAIGVSRISDNGKVIIR 2333
Qy 2102 INGSADTQCHVCAVAGCFHF 2122
Db 2334 LSGTINSQKGTGVAAGVGYOW 2354

```

RESULT 8

```

US-09-268-347-36
; Sequence 36, Application US/09268347
; Patent No. 635182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36

```

```

; LENGTH: 2411
; TYPE: PR
; ORGANISM: Haemophilus influenzae
US-09-268-347-36

Query Match 10.9%; Score 1169.5; DB 4; Length 2411;
Best Local Similarity 22.6%; Pred. No. 7.3e-60;
Matches 618; Conservative 331; Mismatches 855; Indels 925; Gaps 118;

Qy 1 MNHIVYEFKAGTGMVAEYAKSHSGGSCATQVGSVRLTSPARPAALVLYGAT 60
Db 1 MNHIVYEFKAGTGMVAEYAKSHSGGSCATQVGSVRLTSPARPAALVLYGAT 60
Qy 61 LNSAVAGIGISEADGKGANRQDKSIAGIAQ-----ALGOSIAIGDKKIYHNEN 115
Db 61 LNSAVAGIGISEADGKGANRQDKSIAGIAQ-----ALGOSIAIGDKKIYHNEN 115
Qy 53 -----DEDELDVYVTAAPYLSHSDKEGTEGEVENSN 87
Db 53 -----DEDELDVYVTAAPYLSHSDKEGTEGEVENSN 87
Qy 116 NMANIGAKAGNESIAGDVLASGHASIAIGSDDIYKKEVQOISELLPIRQKALN 175
Db 116 NMANIGAKAGNESIAGDVLASGHASIAIGSDDIYKKEVQOISELLPIRQKALN 175
Qy 88 WGIYFNKG-----VLKAGATTLKAG--DNLIKQST----- 117
Db 88 WGIYFNKG-----VLKAGATTLKAG--DNLIKQST----- 117
Qy 176 DIYQADTNLQKRYRTHAQAASHAVGAMSYAKGHSMAFGRATAGTYSVAVGLT--AT 234
Db 176 DIYQADTNLQKRYRTHAQAASHAVGAMSYAKGHSMAFGRATAGTYSVAVGLT--AT 234
Qy 118 -----NASSF-----TSLKKDLTDLT 134
Db 118 -----NASSF-----TSLKKDLTDLT 134
Qy 235 AKAASTIAVSNMAQIGFAATVAG-----GSTQVNLNRGIALGFGSOYLQKNDVYMAN 288
Db 235 AKAASTIAVSNMAQIGFAATVAG-----GSTQVNLNRGIALGFGSOYLQKNDVYMAN 288
Qy 135 SVATEKLSFGANGKQVDTSDANGKLAKTGNVNLN-----GLDSTLDEAVTNGVUS 189
Db 135 SVATEKLSFGANGKQVDTSDANGKLAKTGNVNLN-----GLDSTLDEAVTNGVUS 189
Qy 269 VRAVAPDDNPIDIRYKATKNGATDVESIG--NSNGNDSIRKRIINVGASADTDVAVNA 347
Db 269 VRAVAPDDNPIDIRYKATKNGATDVESIG--NSNGNDSIRKRIINVGASADTDVAVNA 347
Qy 190 SSSFTPNQ--VEKTRATATK-----DVLNAGNNIKG-----ATAGANN-- 226
Db 190 SSSFTPNQ--VEKTRATATK-----DVLNAGNNIKG-----ATAGANN-- 226
Qy 348 QLKEAVRL--ANRQITEKGDSNN--RVEKGLKTLITGGAQTSALTIDHNIQYONG 401
Db 348 QLKEAVRL--ANRQITEKGDSNN--RVEKGLKTLITGGAQTSALTIDHNIQYONG 401
Qy 227 -----ESVDLVAAYNNVEFTGDKNTLDVLTAKENGKTEVEKFTPKTS-----YIKER 275
Db 227 -----ESVDLVAAYNNVEFTGDKNTLDVLTAKENGKTEVEKFTPKTS-----YIKER 275
Qy 402 DGLKVOLEPILTSIKMTTENTLANEKVYVGTGKTRLTDKIGFINDMNGIDSKPYLDKDT 461
Db 402 DGLKVOLEPILTSIKMTTENTLANEKVYVGTGKTRLTDKIGFINDMNGIDSKPYLDKDT 461
Qy 276 DG-----KLTTGKENNDTNKVT--SWTATDN--TDEGGLVATAVAID-- 314
Db 276 DG-----KLTTGKENNDTNKVT--SWTATDN--TDEGGLVATAVAID-- 314
Qy 462 CIHAGGKITRITAGVDDAATYQGLKKNQTAESAALQTFVYKVDKNGN-----DA-- 514
Db 462 CIHAGGKITRITAGVDDAATYQGLKKNQTAESAALQTFVYKVDKNGN-----DA-- 514
Qy 315 AVNKAQMRVKTITANGONGDFATYA--SGTNVTFESDGGTASTYDNTONGITVYKDAK 372
Db 315 AVNKAQMRVKTITANGONGDFATYA--SGTNVTFESDGGTASTYDNTONGITVYKDAK 372
Qy 515 -----NDSKITL-----VGKNNKPDG-----TOVNTLKGENGY 544
Db 515 -----NDSKITL-----VGKNNKPDG-----TOVNTLKGENGY 544
Qy 373 VGDGLKFDSDKKIYADTTALTVTGKVAEIAKEDDKKLVNAGDLVYALGNLSMKAKAEA 432
Db 373 VGDGLKFDSDKKIYADTTALTVTGKVAEIAKEDDKKLVNAGDLVYALGNLSMKAKAEA 432
Qy 545 DVTETNGITYVFGLNQNGGLTVGSLTNDGLSVK-----NTNSKKQIQVAGADGTTFTDIS 600
Db 545 DVTETNGITYVFGLNQNGGLTVGSLTNDGLSVK-----NTNSKKQIQVAGADGTTFTDIS 600
Qy 433 D--DITGCALE--GISKQOEYKAGETVTFKAGKNLKVQDAGANFTYSLODALTGLTITLG 489
Db 433 D--DITGCALE--GISKQOEYKAGETVTFKAGKNLKVQDAGANFTYSLODALTGLTITLG 489
Qy 601 NSKPGAGIENTTTRITRQIGF--ANNFGSLDANKPRLTPTGINAGKELTNVQSAINPAT 658
Db 601 NSKPGAGIENTTTRITRQIGF--ANNFGSLDANKPRLTPTGINAGKELTNVQSAINPAT 658
Qy 490 STTNG--GNDATVINKQGITLTPANGNGTGTNTISTYKQIKRANKNAITNVASGLR--AY 547
Db 490 STTNG--GNDATVINKQGITLTPANGNGTGTNTISTYKQIKRANKNAITNVASGLR--AY 547
Qy 659 NGGOLDPMNRLST-----ANTE--KSGSAATIKDLVNLISQVPLT 635
Db 659 NGGOLDPMNRLST-----ANTE--KSGSAATIKDLVNLISQVPLT 635
Qy 548 DDANFDVNLNNSATDLNRRVEDAYKGLNLNNEKNANKOPLYTDSAAATVGLRKLGVVST 607
Db 548 DDANFDVNLNNSATDLNRRVEDAYKGLNLNNEKNANKOPLYTDSAAATVGLRKLGVVST 607
Qy 696 FAG--DTGNVNTKTLGELLVYKSGKTTADDLTKNNIIGVAVDSTQSLVYKAKTLTSD-- 750
Db 696 FAG--DTGNVNTKTLGELLVYKSGKTTADDLTKNNIIGVAVDSTQSLVYKAKTLTSD-- 750
Qy 751 -----LDAVNTKTLASDKTVVDSGNNTALQNGDLTFSSKONTGATPA----- 793
Db 751 -----LDAVNTKTLASDKTVVDSGNNTALQNGDLTFSSKONTGATPA----- 793
Qy 660 KDGGTILKLYVNQNT-----DNLVIGNNCTIATYKGG--FEVYKGAIDADAGKVTYK 710
Db 660 KDGGTILKLYVNQNT-----DNLVIGNNCTIATYKGG--FEVYKGAIDADAGKVTYK 710
Qy 794 -----TNSKTIYDGLKFTDN--NGIALDGTITYT 821
Db 794 -----TNSKTIYDGLKFTDN--NGIALDGTITYT 821
Qy 711 DATANDAKRVATVAVATAINSAATFVKTEMLTTSIDENPFTDNGKDALAKAGDTLTFK 770
Db 711 DATANDAKRVATVAVATAINSAATFVKTEMLTTSIDENPFTDNGKDALAKAGDTLTFK 770

```

| | | | | | | | | |
|----|------|---|---|-------------------------|------------|-------|------------|-----|
| Qy | 822 | KQVGFAPKOD--- | SLDKSKPYLDK | ---- | DKLVG----- | ----- | EVEITINGIN | 859 |
| Db | 771 | AGKNLKVRRDGNITFDLAKNLEVKYAKVSOTLTIGNTPTGGTTAPKVNITISTADGIN | 830 | | | | | |
| Qy | 860 | ----- | AGKKA1--TGLISMTLND-ATNATTCGHVTOGLJVIDSTDKTPRAASIGDYNINAGF | 908 | | | | |
| Db | 831 | FAKETADYAGSGSKNYVLKGIATTLTEPESAGAKSSHYDL- | NVDATKKSMAASIEDYLRAM | 888 | | | | |
| Qy | 909 | NLKNNGDAKDPYSTDYDFEL----- | ----- | 929 | | | | |
| Db | 889 | NIOGNGNMNDVATYDVTDFNFTDSTGTYTVYVTOKADKGADVKIGAKTSVIKDHNGKLF | 948 | | | | | |
| Qy | 930 | ----- | ----- | 929 | | | | |
| Db | 949 | TGDKLDKANNNGATVSEDDGCKOTGIGLYTAKTYIDAVKSGMRYTGEGATLAEIGATLVNAG | 1000 | | | | | |
| Qy | 930 | ----- | NGNATKAVTYDGKASKVADYVNV-DGTTIHLTGADGNKNOIGVAT | 974 | | | | |
| Db | 1009 | NAEFTVSGTSVFKNGNATPTATVATSKDNGNINVKYDVAVGJLKI----- | GDDKKIVADT | 1066 | | | | |
| Qy | 975 | TTLTFTDK----- | GDKAINSVSGDDKALINAKDILADNLTNLAGEIRINTKCTADTALQ | 1022 | | | | |
| Db | 1063 | TTLVYTGKVSVPAGANSVN----- | NNKLVNAEGLATLNLNLSWTARADK-YADGESE | 1119 | | | | |
| Qy | 1030 | TFQVYKVEENGDDNDADPTITVGDADKATNO----- | VNTLKL----- | KGKG | 1076 | | | |
| Db | 1116 | GETDEYVA----- | GDKTTFKAGKNLKVKSGKEDFTYSLDDTLTGLTSTTIGSTANGRD | 1177 | | | | |
| Qy | 1071 | LDIOTNKGQ-TVTGCIINTOSGLKAGNFTLN-- | NNGLS----- | IKNT | 1109 | | | |
| Db | 1171 | TGTVLNKMGILITTLNGLAAGATDSNONTPTISVTKDGISAGNKEITVWKSALTXYKDTQNT | ----- | 1230 | | | | |
| Qy | 1110 | AGNEOIQYGA----- | DGVKRAKNNNGVAGIDG----- | TTRT | 1144 | | | |
| Db | 1231 | AGATOPPAANTAEVAKODIVLTKRATGATGAAGADAKAPDTTAATVAGDLRGLGWLISAKRT | 1290 | | | | | |
| Qy | 1144 | ----- | RDEIGFAGTNGS----- | LDKSKPH----- | LSKDG- | 1167 | | |
| Db | 1291 | ADETDKEFHAAVKNANNEVEFGVKNATVSAKTDNNNGKHVITIDVAEAKVGGGLEKDDIG | 1350 | | | | | |
| Qy | 1168 | ----- | IMAGKKITTIQSGELTAEONSDNAVTC----- | GKIY----- | DL | 1199 | | |
| Db | 1351 | KIKLKVNTDGNMLLTVDATKAGSAVAGGEENAAVTTTATTAAGTNANERAKVVVKGSGNGAT | 1410 | | | | | |
| Qy | 1200 | KTELENKISSYAKTAQONSLIEFFS-- | VADPEGNNFTVSNPYS-- | SYDTSKTSDVTFP-AG | 125 | | | |
| Db | 1411 | ATEIDKKKVAIVGVVAKAINDAIFVYVEDEDSATIDDSPTDGDANDALKAGDITLILKAG | 1470 | | | | | |
| Qy | 1254 | ENGITTKVNCV-- | VRVGDITQTKGLTTPKLVGNNGK----- | GIYIDSONON | 1300 | | | |
| Db | 1471 | KNLKYKRDGKNITFALANDLSVKSATVSDKLSLGTGNKNVNIITSPDKGLNFAKDSKTGDD | 1530 | | | | | |
| Qy | 1301 | T---TTGSSNTLANVNDKSGSVPTTBEGKIKIKEDKTPRAASIVDYLSAGFINQG----- | 1351 | | | | | |
| Db | 1531 | ANIHNGIASLTLDITLLNSGAT--TNJGCGITDNEKKRPAASVAKDYILNAGMNVRYGVPASA | 1589 | | | | | |
| Qy | 1352 | --NGEAVDVFSTYDTVNFADGNATYAKVYUD-- | DTSKTSKVYUYDVAVDDTTTEVQDKKLQ | 1407 | | | | |
| Db | 1590 | NNQVENIDEVATYDTVDFVSGDKDQTSVYIESKDNRKEV----- | KIG | 1633 | | | | |
| Qy | 1408 | VKTITTLTSGT----- | TGN-KFALSNOAT----- | GDALYKASDIYAHLNTLSGDIQT | 1453 | | | |
| Db | 1634 | AKTSYVIDHNGKLTFTGELDANNNGVYETEDGKDEGNGLTAKAIVADAVKRAMRYKT | 1693 | | | | | |
| Qy | 1454 | AKGASQAN----- | SSAGYVDVADGKKVYIDSDTNKYGAKKNDG-- | TYDJKTTEVAK----- | 1500 | | | |
| Db | 1694 | T-GANGCONDDPATYASGTNVPFADGNGTAEVT----- | KANDGSTITVKNYKVAVGLDKL | 1746 | | | | |
| Qy | 1501 | --DKLVAAQ-- | QTPTDGLAONMNVSVINKEOYVNBANKKOGINEDNAEFVKGLEKAASDNK | 1555 | | | | |
| Db | 1747 | DGDRIYADTITVYLIVADGKVTAPN----- | NGDKK----- | FVDSAGLADALNK | 1788 | | | |

QY 1556 TKNAAYVGDLMNAAGCPLEFAGCGTGTAKKIGETLTIKGGQDTNKTLDNNIGVAGSTD 1615
Db 1789 LSWTA -TAGEGEGEVDPNAPSAGQ-----EYKAGDKVTFPKAG-----DNLTKKOSGK 1834
QY 1616 GFTYVKAKLDLTNLNSV-----NAG-----GTRIDKGVSEYDSSGQA-----KANTPYLSA 1661
Db 1835 DFTYSLKEKELKDLTVSEFKDANGTGSGSESTKITKDGLTTPANGAGAAGANTANTTISVTK 1894
QY 1662 NGDLGKGVLSNNGKGR-----DTPAANYOOL--NEVRLLDLGNAQNDN-----A 1706
Db 1895 DGISAGKKAATVNVSGLKTKKGDDGHTLANGTVADEEKHYDNAVYKDLNLDEKADNNEPTVA 1954
QY 1707 DGNQVNIAD-----TKDPNSGS-----SSNRTYIRAGT-----VLGG-- 1739
Db 1955 DNTAATVGDRLGQWVTSADKTTGEPNQEVNAQVRYANNEVKFPKSGNINVSCKTLNSTRV 2014
QY 1740 -----KGNDEPKATG-----VOVG-----YKDGNAQDLSNVWYKTK-- 1776
Db 2015 IYTELAKGEVVKNEFEYVKNADSETNLVYKGVGMYYSKEDIDATSPKPMTC--KTEYKY 2071
QY 1777 -----DGSKKALATYNAAGQVYTLNPNPAEALDRINEQIRPEFHVNDGQNEPYQ 1827
Db 2072 VENGKVVYSAKSGKTEVYLTINKSG--YVYGN--QVADALAKSGFEL----- 2113
QY 1828 GRNGIDSSAGKHSVAIGFOAK-----ADGEAAVAIGRQTOAGNOSTAIGDMAQ 1876
Db 2114 ---GLADAAAEKAFESADKQLSKDKAETVNAHDKVFPANGCLNTKV---SAATVESTD 2167
QY 1877 ATGDO-----SIATGTNVVAAGKHSAGIGDPSTYKADNASTSVGNNNOFTDATQTDVF 1928
Db 2168 ANDDKYTTTFVKTDVDELPLTQIY---NTDANGNKIYKADGKV-----YELNAD 2213
QY 1929 GVGNNITVYESNVALGSNSAISAGTHAGTOAKKSDGTACTTTAGATGTGKFPAGQTA 1988
Db 2214 GTASNKEVLTGNVYDANGK-----YKVTENGADKKWYTTNADGAAADTKGEYSN 2262
QY 1969 GAVSVG-----ASGAERRIQNVAAAGEVSATSTDAVNGSQTYKATQCIAN---A 2033
Db 2263 DKYSTDEKHVRLDPNNOSNGKGVVIDVANGEISATSTDAINGSQLYAAVAKGVTNLAGQ 2322
QY 2034 TNELDRHIOHNEKNKANAAGISSANAMAMSMPOAYTPGSMYGTGIATNNGOGAVVAGLSKLS 2093
Db 2323 VNNLEKVKVYKRRADAGTASALAAQDLQATMPGKSMVALIAGSSVYQONGLAIGVSRIS 2382
QY 2094 DNGQWFKINGSADTOGHVGAAGVGFHF 2122
Db 2383 DNGKVIIRLSGTTNSQKGTGVAAVGQYOW 2411
RESULT 9
US-09-268-347-28
; Sequence: 28, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 1104
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-28
Query Match 9.5%; Score 1018.5; DB 4; Length 1104;
Best Local Similarity 27.8%; Pred. No. 1.7e-51;
Matches 368; Conservative 165; Mismatches 399; Indels 393; Gaps 53
QY 943 KASKVAIVDVNDGTTIHLTGADGNKNOIGVKTTLTKTKDAKGDKAIFNSVNSGDD----- 997

Db 28 KCAATYAAVAVLATATLSAT-AEANNN-----TSVTNGLNAYGDTNFTNNTNSIADLEKHV 81
 QY 998 ----KALINAKD-----IADNLTLAGELR-----NFKGTADTALQTFQYKVK 1037
 Db 82 QDAYKGLINLEKDTNKSSEFLVADNMTAATVGNLRLGLWVLSKRGTRNE--KSYQVKAD 139
 QY 1038 ENGDDNDADDTITVGNKAKTNOVNTLKLKGNGLDIOITNKGDTITPINTQSLKAGNNT 1097
 Db 140 E-----VLTGSGAATVS-----SSSKDGKHTTITVTKGSFAEVKT 176
 QY 1098 TLNNGLSIKNTAGNEOIOYAGADGVKFAKYNNGVAGIGTTRITRDEIGFAGTNGSLD 1157
 Db 177 DATGG-----QVNADEKGVKADEN--GADV-----202
 QY 1158 KSKPHLSKDGINAGGKTTNIOGSEIAONSNDVATGKIYDLKTELENKISSTAKTAONS 1217
 Db 203 -----KKAATVK--DVAKAIINDATFVKVESTDDDIENGAAGKNETTDOA 245
 QY 1218 LHESVADDEGNNFTV---SNPYSSYDTSKTSVITFAGENGITTVKNGKGVAVGIDQTK 1274
 Db 246 L-----KAGDTLTLKAGKMLKAKLDQNGKSVTFALAKDLDTVSARVSDKLSIGKDTNK 298
 QY 1275 -----GLTTPKLTGYNNGNGKGIYDSONGONTITGLSTLTANTNDKGSVRETEQCK 1326
 Db 299 VDTISDANGKLAK--TGNGNG-----QNGVHLNGLASTL---TDTITGMTQASN 345
 QY 1337 ITKDEKTRAASIVDLVLSAGFNLONGEAVDVSTYDTPVNFADGNATTAKYVYDTSKTS 1386
 Db 346 GVAVQNNHRAASVADVNLNAGMNIQNGASVDFVNAVDTVDFVNGTNNVNTVTTDTHAKKT 405
 QY 1387 KVVVD-----VNVDDTITIEVK-----DKKLK-----VTTTITLSTGTG 1419
 Db 406 TVRVDVYTGLPVOYVTEDEKTVKVKVKNYKYEAKODGSADMDKVENGELAKTKVLSVAG 465
 QY 1420 ANKFPALSNQATGDLAVKASDIVA--HLNLTLSGDIOTAKKASQANSSAGYVADADNGYV-- 1475
 Db 466 QNPVKISNVAG---TEENDAVSFQOLKALO-EKQVTLTASNAANAANGN--DADGKATOT 520
 QY 1476 -----YDSTDNKKYQAK--ND-----GTVDKTRKFAKDKLVAQAOTPDGTIAQM 1517
 Db 521 LNNGLNFKSTDELINIKIENDTVTFTPKGSV--QVGEDEKATIQNGKRTTGTGLVEAS 579
 QY 1518 NVKSVINKEOVNANKOGIEMEDNAFVKGLEKASDNKTKNAATVDDIADNAVQDTPLFA 1577
 Db 580 EL-----VESLNKLGKMGVVDKGS-----GELDASNETLVKS 613
 QY 1578 GDTGTTAKLGETLTKIGGQDTNKLTDNINIGVAGTDFVFLAKDLTNSV-----1631
 Db 614 GDKYTL--KAEENLKVK--QDGTN-----FTYALKDELITGVKSVFEKDTA 654
 QY 1632 ---NAGGTRIDKGVSEVDSG---QAKANTPVLNANGIDLGKYSINVGKGTCTDA 1683
 Db 655 NGSNGASTRTKIDGLTITSANGANGAATDADKIKVASDGSISAKNKAIVKNVSGLKKEGD 714
 QY 1684 ANVOQLNVRULL-----GLGNAGNDADGNOVNIAIDIKRDPNGSSSNTFTVAKG 1734
 Db 715 ANFNLJTSSANLTKQYDAYKGLTNDEKAGDKOTLVAD---NR-----AA 759
 QY 1735 TVLGKGNNDTEKLATGQVGVVDKGNANGDLNV--WVKTOKDGSKALLATYVNAAGOT 1793
 Db 760 TV-----GDLRLGLNV--ISAOKTIGELKKEVNA-----786
 QY 1794 NYLNNPAEALDRINEGIRFFHVNDGNOBEVVOGRNGIDS---ASGKSHVAIGFOAKA 1850
 Db 787 -----QVNAANEVFKKSG--NGIHVSGKTVNGREIT--FELAK 821
 QY 1851 DGEAAVATGROTQAGNOSIATIGDNAOATGDSIAIGTGNVYAGKHSAGIGPSTV--KAD 1908
 Db 822 DENA-----IAFGYSGKALRDNVTVAIGTGNVNAEKSAGFGPNTIEDKAG 867
 QY 1909 NSYVGNNOFTDATOTDFVFGVNN-----TTV-----T 1937

Db 868 GSYAFGNDRKIT-SKNTPFLNCGNNAKYKANGVDVDETIVTKDKDEKETTIVPRALCAT 926
 QY 1938 ESNVVALGNSNSASAGTACGTOAKKSDGAGTTTGAATGTYKGFAGQAVAVSVGASG 1997
 Db 927 VENSVALGNK--STATKDKGKMLKSDGTAGNTTATAGTGVNGFAGATAGAVSVGASG 983
 QY 1998 AERRIONVAAGEVASTDAVNGSOLYKATOGIANTNELDRHIONEKNANAGISSAMA 2057
 Db 984 FERRIONVAAGEISATSTALNSOLYAAKGYTN---LAQVKNVGRADAGTASALA 1039
 QY 2058 MASPMQATYIGRSMVYGTATNNGOGAVAVGLSKSDNGQWFEKINGSDATGHWGAAGV 2117
 Db 1040 ASQLPQASPMGRKSMVSIAGSSVQOGGLAIGVSRISDNKVIIRLSGTNSQKGTGVAG 1099
 QY 2118 ACFHF 2122
 Db 1100 VGYOW 1104

RESULT 10
 US-09-268-347-34
 ; Sequence 34, Application US/09268347
 ; Patent No. 6335182
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, "Sheena M."
 ; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
 ; FILE REFERENCE: 1038-860
 ; CURRENT APPLICATION NUMBER: US/09/268,347
 ; CURRENT FILING DATE: 1999-03-16
 ; NUMBER OF SEQ ID NOS: 54
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 34
 ; LENGTH: 1104
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 US-09-268-347-34

Query Match 9.5%; Score 1018.5; DB 4; Length 1104;
 Best Local Similarity 27.8%; Pred. No. 1.7e-51;
 Matches 368; Conservative 165; Mismatches 399; Indels 393; Gaps 53;

Db 943 KASKVAVDVNDGTTIHLTGADGNKNOIGVTTTLTKTDKAGKAINFSVSGD-----997
 Db 28 KCAATYAAVAVLATATLSAT-AEANNN-----TSVTNGLNAYGDTNFTNNTNSIADLEKHV 81
 QY 998 ----KALINAKD-----IADNLTLAGELR-----NFKGTADTALQTFQYKVK 1037
 Db 82 QDAYKGLINLEKDTNKSSEFLVADNMTAATVGNLRLGLWVLSKRGTRNE--KSYQVKAD 139
 QY 1038 ENGDDNDADDTITVGNKAKTNOVNTLKLKGNGLDIOITNKGDTITPINTQSLKAGNNT 1097
 Db 140 E-----VLTGSGAATVS-----SSSKDGKHTTITVTKGSFAEVKT 176
 QY 1098 TLNNGLSIKNTAGNEOIOYAGADGVKFAKYNNGVAGIGTTRITRDEIGFAGTNGSLD 1157
 Db 177 DATGG-----QVNADEKGVKADEN--GADV-----202
 QY 1158 KSKPHLSKDGINAGGKTTNIOGSEIAONSNDVATGKIYDLKTELENKISSTAKTAONS 1217
 Db 203 -----KKAATVK--DVAKAIINDATFVKVESTDDDIENGAAGKNETTDOA 245
 QY 1218 LHESVADDEGNNFTV---SNPYSSYDTSKTSVITFAGENGITTVKNGKGVAVGIDQTK 1274
 Db 246 L-----KAGDTLTLKAGKMLKAKLDQNGKSVTFALAKDLDTVSARVSDKLSIGKDTNK 298
 QY 1275 -----GLTTPKLTGYNNGNGKGIYDSONGONTITGLSTLTANTNDKGSVRETEQCK 1326
 Db 299 VDTISDANGKLAK--TGNGNG-----QNGVHLNGLASTL---TDTITGMTQASN 345
 QY 1337 ITKDEKTRAASIVDLVLSAGFNLONGEAVDVSTYDTPVNFADGNATTAKYVYDTSKTS 1386
 Db 346 GVAVQNNHRAASVADVNLNAGMNIQNGASVDFVNAVDTVDFVNGTNNVNTVTTDTHAKKT 405


```

QY 1387 KVVYD-----VAVDDTTIEVK-----DKRLG-----YKTTLTSTGCG 1419
Db 406 TVAVDTYGLPVQYVTEDEGKYVAVDKKYYEAKQDSADMDKVENELAKTKVLYSASG 465
QY 1420 ANKFAISNOATGALVAKASDIYA--HLNLTSGDIQAKASQANSAGYVADANKVI-- 1475
Db 466 QNFKISNVAEG--TEENDAVSFQOLKALQ--EKQVTLTASNAAYANGCN--DADGCKATQT 520
QY 1476 -----YDSTDNKYYQAK--ND-----CTVDKTEVAKDKRIVAOQPPDGLAOM 1517
Db 521 LNNGLNFKFESTDGLLNKIKVENDVTFPPKGSV--QVGEDGKATLQNGTKTTDGLVEAS 579
QY 1518 NVKSVINKEDOVNANKKQGINEDNAFVKLEKASDNKTKNAAVYDGLNAAQPLTFA 1577
Db 580 EL-----VESLNLKMKRKYGVDRKGS-----GELDGASNEPLVVS 613
QY 1578 GDTGTAKKIGETLTIKGGQTDNKLTDNNIGVAGTDFYVKLANDTLNLSV----- 1631
Db 614 GDRVTL--KAGENMLKYK--ODGTN-----FTYALKDELIGKVSVEFKDIA 654
QY 1632 ---NAGGTRIDDDGVSFVDSG-----QAKANTPVLSANGLDGKVISNVGKGTDTDA 1683
Db 655 NSNGASTFKITKDGTLTTSANGANGAATDADIKVASDGSISANKAVNVSGIKRFGD 714
QY 1684 ANVOQLENEVENLT-----GLGNAGNDNADGNQVNIADIKKDPNSSSSNRVYIKAG 1734
Db 715 ANGNPLTSSADNLTQYDDAYKGLTLNDEKADKOTLTVAD-----NT-----AA 759
QY 1735 TVYLGKGNNDTEKLATGVGVGVVDKNGANDLSNV--WKTQKDSKALLATYNAAGT 1793
Db 760 TV-----GDLRLGLGV--ISADKTTGELNKEYNA----- 786
QY 1794 NYLTNNPAEADIRINEGIRFEHVNDOEPVVOGRNGIDSS--ASGHSVAIGFOKA 1850
Db 787 -----QVRNANEVKKESG--NGIHSVSKTYVNGRREIT--FELAK 821
QY 1851 DGEAAVAIGHQTOAGNOSIAIGDMAQATGDOSIAIGTVGVVAGKHSIGADPSTV--KAD 1908
Db 822 DENA-----IAFGYSKALRDNVTALIGTVVNAEKSGANGDNYIEDKAG 867
QY 1909 NSISVGNNOFTDATQTDYFVGNN-----TIV-----T 1937
Db 868 GYAFANDNMT--SKNTPVLGNGVNAKRYKANGDVTEFVTKDKCKETTJVVPKALGAT 926
QY 1938 ESNSVALGNSATSAGTHAGTOAKSSDGTAGTTTATGATGVKGFAGOTAVGASVGSAG 1997
Db 927 VENSIVLGNK-----STATKOKGKMLKSDGTAGTTTATGTTVNGFRGATAHGAVSVGSAG 983
QY 1998 AERRIONVAGEVSATSTDAVNGSQLYKATOGIATNATNELDHRHONENKANAGISSAMA 2057
Db 984 EERRIONVAGEISATSTDAINGSOLYAVAKGYTN---LAGOVNKKVGRADAGTASALA 1039
QY 2058 MASMPAYITPGRSMYTGATIHNGCAVAVGSLKSDNQWFKINGSADTQGHGVAAG 2117
Db 1040 ASOLPQASMPGRSMYSIAGSYOGONGIAIGVSRISDNKGVIIRLSTGNTSGKTGVAG 1099
QY 2118 AGEHF 2122
Db 1100 VGTQW 1104

```

RESULT 11

US-09-268-347-30

Sequence 30, Application US/09268347

Patent No. 6335182

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS

FILE REFERENCE: 1038-860

CURRENT APPLICATION NUMBER: US/09/268,347

CURRENT FILING DATE: 1999-03-16

NUMBER OF SEQ ID NOS: 54

SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 30
 : LENGTH: 1004
 : TYPE: PRT
 : ORGANISM: Haemophilus influenzae
 : US-09-268-347-30

Query Match
 Best-Local Similarity: 9.1%; Score 974; DB 4; Length 1004;
 Matches: 337; Conservative 146; Mismatches 330; Indels 382; Gaps 45;

```

QY 1073 IQTNKDGTVTFGINTQSLKAGNNTLNNGLSIKTNAGNEFOIGVADGVFAKVNNGV 1132
Db 43 LSTVQATTTGGTISTNKLKAVGT--NNPF--MAAGN----- 77
QY 1133 GAGIDGTRITRDEIGPAGT--NSLDSKPLSLKSDGINAGKKTITNIOGELIANSNDAY 1191
Db 78 ---SATDLARO--FDGAYDGLNLNLEKDNKN-----LVTDDKAA 113
QY 1192 TGGRKYDKTELEKISSTAKTAQNSLHEFSVADQGNFTVSNPSSYDSTKSD---V 1248
Db 114 TVGNLRKLGWVLSKNGRNEKSOQVKADEVLEFGDGVTV-----TSSENGKHT 165
QY 1249 ITFAGENGITTRKVNKGVVRVIGIDQTKGLTTPKLTGVNN-----NGKGIYDSQNG 1298
Db 166 VTFLEKDLANK-----NATVSDKLSLGNANGKRVDTSPDNGLKFAKPTNG 212
QY 1299 QN---TTIGLSNLTANVTNDKGSVTTTEOGKTIIDEDEKTRAASIVDYLSAGFNLONGEA 1355
Db 213 QNGNVHNLGIASTLTDLT--GTRKSATNG--VDVQHNRAASVADYDLNMGWNTIOGNGAS 268
QY 1356 VDEVSATDVTNPAADGNATTAKTATVYDDTSKTVYDVNVDDTTEVVDKLGVTTLTUS 1415
Db 269 VDFVATYTFDVPFVGNLNTNVAVT--TDIAHNKKT--VRVDTGLPVQ----- 312
QY 1416 TGTGANKFALSNQATGALVAKASDIYVAHLNLTSGDIQAKASQANSAGYVADANKVI 1475
Db 313 -----YVTDEGETV 322
QY 1476 YDSTDNKKYYQAKNDGYDKTEVAKDKL-----VAQKQTPDGTIAQNNVK 1520
Db 323 ---KYGNEEYFEAKQDSADMDKRVENGKLAKTKVLYSANGTNPVKNVADGT--ENTD 377
QY 1521 SVINEQVANDANKKOGINEDNAFV-----KGLG-----KAA 1551
Db 378 AVSEFQOLKALDDQVYTLASNAAYANGSGDADGAGIOTLSGLNLFKFKSTIDGLNINAE 437
QY 1552 SDNKT--KNAAVTVGD-----LNAVAQPLTTPAGDTGT 1583
Db 438 NDVTFTFPKKSQVQGDGKATIDGAKTTGLTGLVEASELVSLNKLGMKVGOTGDTGVT 497
QY 1564 AKKLGETLTITGQTDNKLTDNNIGVYAGTGTIVLALNDTLNLSV-----NAG 1634
Db 498 DGHATDYL--VRSQKVTYLRKAGDNLKVKOEFTN--FTYALKBELVDYKVERKDRANGANGA 555
QY 1635 GTKIDDKGVSEVDSGQA-----KANTPVLSANGLDGKVIKISVWGGTQDTDAANYQOL 1689
Db 556 SKTIKQDGLITTPANGAGACANTANTISYTKDISAGNKRAVKNVYSGLKFFGANGDPL 615
QY 1690 NEVRNLGLGNAGN---DNADGNQVNIADIKKDPNSSSSNFTVTKACTVLOGKGNND 1744
Db 616 T-----SSADNLTQKYDNAKGLNLDKESKQKQTPVADNT--AAV----- 656
QY 1745 TEKLATGVGVGVVDKNGANDLSNV--WKTQKDSKALLATYNAAGQNTYLTNNPAEA 1803
Db 657 -----GDLRLGLGV--ISADKTTGELNKEYNA----- 681
QY 1804 IDRIINEGIRFEHVNDOEPVVOGRNGIDSSASAG---HSAVIGQAKADGPAVAIG 1859
Db 682 -----QVRNANEVKKESG--NGI--NSGKGLDNKGTREITTELAKDEKA----- 721
QY 1860 ROTQAGNOSIAIGDMAQATGDOSIAIGTVGVVAGKHSIGADPSTV--KADNSYSVGNNN 1917

```


Db 722 -----IAFGSGSKALRDNTVAIGTGNVVAEKGSGAGDPNTIEDKAGSGYAGCNDN 772
 QY 1918 QPDTATQTDVFGVGNNT-----IVTESNSV-----AG-----SNSAIS 1951
 Db 773 RI-----ISKNTFVLGNSVNAKRDANGVNLTEKEVKGDKAKTVYVPAALGETVEVNSVYL 829
 QY 1952 AGTHAGTQAK-----KSDGTAGTTTACATGTVKGFAGOTAVGAVSGASGAEERRIQNVAA 2007
 Db 830 GMASTAKDKGKNLKSQGTGNTTATGATCTVNGFAGATLHGAVSGASGAEERRIQNVAA 889
 QY 2008 GEVASTATDAVNSQILYKATQIGANATNELDHRHONENKANAGISSAMAMASPOAYIP 2067
 Db 890 GEISATSTDAINGSQLYAAVAKGVTN-----LAGOVNKGKRRADAGTASALASOLPOASMP 945
 QY 2068 GRSVNTGCIATHNGOGAVNGSLKSLSDNGOWEKRINSAPTOGHVGAAGVAFHF 2122
 Db 946 GKSWSVSIAGSSYOGSGLAIGVSRISDNGKVIIRLSGTTNSOGKTVGAAGVGYOM 1000
 RESULT 12
 US-09-268-347-24
 ; Sequence 24, Application US/09268347
 ; Patent No. 6335182
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
 ; FILE REFERENCE: 1038-860
 ; CURRENT APPLICATION NUMBER: US/09/268,347
 ; NUMBER OF SEQ ID NOS: 54
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 24
 ; LENGTH: 1002
 ; TYPE: PRP
 ; ORGANISM: Haemophilus influenzae
 ; US-09-268-347-24
 Query Match 8.88: Score 944.5; Db: 4; Length 1002;
 Best Local Similarity 29.5%; Pred. No. 3e-47;
 Matches 338; Conservative 134; Mismatches 342; Indels 333; Gaps 46;
 QY 1125 AKVNVGVAGIDGTFTRTRREIGFAGTNGSLDKSRPHLSKIDGKNAGCKITINQOSEIA 1184
 Db 40 ATVLSATVQASAGST-----GTN-----SLNVYGNKNSNFSN---A 73
 QY 1185 QNS-----NDAVTGKGIYDLKTELENKISSTAKTAQNSIHEFSVADEQ-----GN----- 1229
 Db 74 NSTADLNKQNDY-----YDGLLNLNEKGTDKSK-----FLVADETTAIVGNLAKL 120
 QY 1230 NFIYVNSPYSYDTS-----KTSQVITPAGEBG-----ITTKVNGVVRVYDIDOTGLTTPKLTVG 1284
 Db 121 GVVVSTKNSKSTEEESNOVKQADEVLEEGKDGVTYVTSKSENGHVTFLPALDLANKNATVS 180
 QY 1285 NN-----NGKGIIVDSQ-----NGONTTGLSNLTANTNKGSRTEEQKII 1328
 Db 181 DKLSLGAAGKVVDTISDANGIKFAKQGTNGNGNVHLNGLASTIDDPVGGKTAHLTKRI 240
 QY 1329 KDEDETRASTIVDVLSAGFNLO-----GNGEADVFSTVDVVFAGNATTAKTAVTYDDT 1382
 Db 241 SDTERNRASVGDVLNMGWNRGAKKTIGGYDNDVDFVSTYPTVFAGAGANANVSYTTDDN 300
 QY 1383 SKTSKVYVDVNDVDTLEVK-----DKKLG-----VKTT 1411
 Db 301 KKT-----VAVDVTGLPVOYVTEDESKTVVGVNEVEYEAQODGSADMDKVEYNGKLAKTK 355
 QY 1412 TLITSTGANKFALSNQATGALAKADIYA-----HLNLTSDIQIARAKASQANSAGVYDA 1469
 Db 356 VKLVSANGTNPVKLISNADG-----TEDDVAVFQKQKALQ-----DKOVTLSASNAVYAGG-SDA 410
 QY 1470 DGNKVI-----YDSTDNKRYQAKNDG-----TVDKTK-----EVADKDLVAAQOTPDGT 1513

Db 411 DQKATQTLGNDLNFKFKSTDELLNIKAGDVTFTPKGSGVGYGD-----GKATIDG- 466
 QY 1514 LAOMNVKSVYINKEOVYNDAR-----KQGINEDNAFVNGLEKASDNTKNAAYTVGDLNVA 1570
 Db 467 -AKTTTGLVASELYDSLKLKGMKVGKDG-----GATD----- 501
 QY 1571 QPPLTFAGDTGTTAKKLGELTITKGGQDTNKLITDNNIGVAVAGTDFYVKLAKDLTJNLS 1630
 Db 502 -----GHHTDILVSGSKVTLKAG-----DNLRKREGTNPFTVYLADELTLGVS 545
 QY 1631 V-----NAGTKTIDKVSFV-----DSSQA-----KANTPVLNANGLDLGGKVLNNGK 1676
 Db 546 VEKKTIDENGANGASTKITKDGGLTTPANDANGAATDADKIKVADSDISAGKAKVNVVS 605
 QY 1677 GTKDTDAVVOOLEVNRNLGLGNAGN-----DNADGQNVNADIKKDPNSGSSSNRFTY 1731
 Db 606 GLKFGDANFNPLT-----SSADNLTQYDINAAYKGLTNDESKSKQOTPTVADNT-- 655
 QY 1732 KAGTVLGGKGNNDTEKLTAGVGVGVKDGDNANGDLSNV-WV-----KTQKDSKKALLATYN 1788
 Db 656 -AATV-----GDLRGLGWISADKKTGTGESEKYSAGVR 686
 QY 1789 AAGQTVYLTNNPPEALIDRINEGCIREFHYNDGQEPVYVGRNGIDSSAGSKVAIGQIA 1848
 Db 687 MNEVEFKSGNGCINVSCKTIDNGTR-----BITTEL 717
 QY 1849 KADGEAAVAIAGRTQAGNOSTIAGNOAATGOSTIAGTGNVAGKHSAGIDPSTV--K 1906
 Db 718 AKDNA-----IAFGSGSKALRDNTVAIGTGNVVAEKGSGAGDPNTIEDK 763
 QY 1907 ADSVSVGNNOFTDATQTDVFGVGNNT-----ITV----- 1936
 Db 764 AGGSYAFQNDNRIT--SKNFFVLGNGVNAKYKANGVDVETVYVKDKGKETTIVYPKALG 822
 QY 1937 -TESNSVALGSSNSAISAGTHAGTQAKKSDGTGTTTACATGTVKGFAGOTAVGAVSVA 1995
 Db 823 ATVENSVYLGK-----STAKDKGKNLKSQGTGNTTATGCTVNGFAGATLHGAVSVA 879
 QY 1996 SGAEERIONVAAGEVASTDPAVNSQILYKATQIGANATNELDHRHONENKANAGISSA 2055
 Db 880 SGERRIONVAAGEISATSTDAINGSQLYAAVAKGVTN-----LAGOVNKGKRRADAGTASA 935
 QY 2056 MAASMPQATIPGRSVNTGCIATHNGOGAVNGSLKSLSDNGOWEKRINSAPTOGHVGA 2115
 Db 936 LAASQLPQASMSKSNVSIAGSSYOGSGLAIGVSRISDNGKVIIRLSGTTNSOGKTVGA 995
 QY 2116 VGAGFHF 2122
 Db 996 AGVGYOM 1002
 RESULT 13
 US-08-409-995-4
 ; Sequence 4, Application US/08409995
 ; Patent No. 5646259
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen I.
 ; APPLICANT: St. Geme III, Joseph W.
 ; TITLE OF INVENTION: Haemophilus Adhesion Proteins
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
 ; STREET: Four Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/409,995
 FILING DATE: 24-MAR-1995
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Silva, Robin M.
 REGISTRATION NUMBER: 38,304
 REFERENCE NUMBER: A-61053/REF
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELEX: 910 27299
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1912 amino acids
 TYPE: amino acid
 STRANDEDNESS: double
 TOPOLOGY: unknown
 US-08-409-995-4

Query Match 7.9%; Score 845; DB 1; Length 1912;
 Best Local Similarity 22.2%; Pred. No. 4.4e-41;
 Matches 478; Conservative 254; Mismatches 646; Indels 772; Gaps 96;

QY 1 MNHKKVFNKATGTFMAVAEYAKSHSTGGSCATGQVGS-----VRLSFARIALAVL 55
 DB 1 MKKIFVNIWNTQTWVYVSELTRTH-----TKRLNRDQVPLATLLFATVQANMT 51
 QY 56 VIGATLNGSAVAGIGISEADGGKGANARQKSAIGDIAQ-----ALGSGSIAIGDKTI 110
 DB 52 -----DEDELDVVPVTRAPVLSHSPKKEGKEV 81
 QY 111 VHSNNNNANIGAKAGNESIAIGDVLASGHSIAIGSDDLTKKEFYQOISSELLPIRG 170
 DB 82 TENSNNKGIYFDKNG-----VLKAGATTLKAG-DNLKXKQXDE----- 118
 QY 171 OKALNDIYQADINTLOKRTTHAOGHASTAVGAMSYGKHFESNAPTRATAGTYSLAVG 230
 DB 119 -----XTNASSF-----TYSLKLD 132
 QY 231 LT-ATKKAASSTAVGSMQAIQFAATANG-----GSTQVNNKRIALGFGSOVIQKND 283
 DB 133 LVDLTSVATEKLSFGANGKVDITSDANGKLAKTNGNVHLN-----GLDSTLPDAVTN 187
 QY 284 VNAANVRAAPDDNQPIDNRKATKFGKATGVESIG-NSNGNDSIRKTIINAGSADTD 342
 DB 188 TGVLSSTSTPMD-----VEKTRATYK-----DVLNAGNMNIK-----AKTA 225
 QY 343 AVNAVLKAVRL--ANROITEFGDSDNNRREKGLKTLITIGAGOTSALT-DHNIGVQ 399
 DB 226 GGNV-----ESVDLSAVNNVEFITGDKMT-----LDVVLLAKENKKTTEVKTPTTSVIK 276
 QY 400 NODGLKVLAEITLSLKAVTTENLANEKYVYKTRITLTDKIGFNDMNGIDESPYYDK 459
 DB 277 EKDG-----KLTGKENNDTNKVT-----SNTATDN-----DDEONGGLVTAKAVID- 317
 QY 460 DTGIIAGGKITKLTVAGVDDDAATYGOLKKVNOTAESALOTFFVKYKVDKNN-----D 513
 DB 318 --AVNKAAGRVTATTANONGNDFAIVA--SGTNVTFESGDDGTTSVTVTDTONGCITVAYD 373
 QY 514 A-----NDSKIITVGNKNNKPGDQVNTL-----KLKGENGVDTT----- 548
 DB 374 AKVGDGLKFPDSKKTIVADTTLTVGKVAELAKEDKKLLVNAADDLVTAIGNLSMKAKA 433
 QY 549 --ETNGYTFGLNONGGLVNGSTLANONGLSYK-----NTNSKQIQVAGADGTTTFDISNS 602
 DB 434 EADDTGALAE-GISKDOEKYKAGETVTFKAGKNNKLVKQDGANFTYSIQDALTGITSTLTIGT 492
 QY 603 KKGAGIENTRTTRDGIGF--ANNNGSLDANKPRLPTGIANAGKELNNVOSAIIPATNG 660
 DB 493 TNG-GNDKATVINKDGLITTPAGNGTGITNTISTYKGIKAGKAITNVASGLA-AIDD 550
 QY 661 GOLDFMNLST-----ANTE--KSGSAATIKDLNLSQVPLTFA 697

DB 551 ANFDVLNNSATDLNRHVEDAVKGLLNENKANNOPLVDTSTAFATVDLKLGMVSTKN 610
 QY 698 G-DTGPVNTKKGLEILKVGKGTADTLTKNNICVAVSDNSITVKYLAITLSDL----- 751
 DB 611 GTRKEESNOVQADEVLTTGGAAATVTSKSENG-----KHTTVSVAETKADCGLEKD 662
 QY 752 -DAVNTKTLTA-SDKVTVDSGNNNTAKLQNGDLTFSSKONTGATPA----- 793
 DB 663 GDTIKLAVDNDNTDNLTVGNGNGTAVTKG--FEYTKATGATDADRKVTVKATANDAD 719
 QY 794 -----TNSKTIQVGLKFTDN--NGIALDGTYYITRKQVFAK 829
 DB 720 KKVATVDVATAINSAATFVKTEMLTTSIDENPTDNGKDALKAGDTLFFKAGKMLKYA 779
 QY 830 QDG-----SLDKSPYLDK-----DKLYG-----EVEITTNQIN----- 859
 DB 780 RDGKNITFDLAKNLEVTAVASDTLITIGNTPTGGITTATPKVNTSTADGLNFAKETADA 839
 QY 860 AGGKAI--TGLSNTLTD-ATNATTGHTVQGIYDSTDKTFRASIGDVLNAGFNLKNNGDA 916
 DB 840 SGKNVVLKGIATTLTPRSAGAKSSHVDL--NVDATKSNAASTIEDVLRAGWNIQNGNN 897
 QY 917 KDFVSTYDYDEI----- 929
 DB 898 VDVYATVDYVNFETDSTGTTTVTVTKADGKADVKIGAKTSYIKDHNGKLTGDKLDA 957
 QY 930 ----- 929
 DB 958 NGATVSEDDGKDTGLVAKTVIDAVNKSQWRYTEGATETGATVAVNAGNAETVSG 1017
 QY 930 -----NGNATTAKVTVYDGASKAVYADVNV-DGTTIHLTGADGNKQIGVKTTLTKTDA 982
 DB 1018 TSVNFKNGNATFVYSGDNGNINVKYDVNGDGLKI--GDDKIYADTTLTVTGG 1071
 QY 983 K--GDKAINESVNSGDDKALINAKDIADNINLTAGEIR-----NTKGTADTALO- 1029
 DB 1072 KVSVPAGANSVN-----NNKKLVNAGLATALNNLSWTAKADKVDAGESEGETDEYKA 1125
 QY 1030 -----TFQV--KKYKEN----- 1039
 DB 1126 GDKVTFKAGKNNKVKQSEKDFYSLQDTLGLTSLGTLGANGRNDGTVINKGLITTL 1185
 QY 1040 -----GDDNDADDTITVGD-----AKTN 1058
 DB 1186 ANGAAGTSDASNGTISVTYDGISAGKKEITNVKSALKTYPKQONTADERTQDEPFAAVK 1245
 QY 1059 QVNTLKLKNGKIDIQTKDG--TVTFGINFQSGIKAGNNTLLNNG--LSIKNTAG 1111
 DB 1246 NANEVEFVGKAGTAVSAKTNNNGKHVTIDV--AEAKVGDGLEKPTDGKIKLKVNTDG 1302
 QY 1112 MEQIQVAGADVKKFAVNVNGVGA-----GIDGTRITREIIFACTNGS--LQSKRP 1161
 DB 1303 NNLLTV--DAIKGASVAKGFEANVTDTATTAOGTANANERKVVVKKSGNGTATETEK-- 1357
 QY 1162 HLSKDIGINAGKKTNTIOSEIANSNDVATYGGKIYDLKLELENKISSTAKTAQNSLHEF 1221
 DB 1358 -----KVAATV--GDVAKAINDAATFYKV-----ENDSATI-----DD 1389
 QY 1222 SVADEGNNFTVSNPYSSYDTSKTSVITE-AGEENITTKVNGV--VRVGIDQTKGLT 1277
 DB 1390 SPTDDGAN-----DALKAXDITLTKAKKNNKLVKRDGKNITFALANDLSVSATV 1438
 QY 1278 TPKTLTVGNNGK-----GIVDSQNGVNT--ITGLSNTLVANVNDKSVYTTQD 1324
 DB 1439 GSKLSTGTNGKNNVITSDTKGLNFAKSKTGDDANIHNLNIASTLITDILNSAT-TNIG 1497
 QY 1325 GKTIKDEKTRASIVVLSAGFNLOG-----NGEAVDFSVTYVAFADGNAFTAK 1376
 DB 1498 GNGITDNEKKAASVRYVLAAGNVRGVRPASANNQVENIDFAATYDLYDFVSGDDDTIS 1557
 QY 1377 VTVD--DTSKTSKVVYDVNVDDTTIEVKDKLGVKTTTLTSTG--TGAN-KFALSNOA 1429

Db 1558 VYTESKDNKRTVEY-----KICAKTSVIRKDHNGKLFTEGKELKDANNNGV 1601
 QY 1430 T-----GDALVYASDIYVAHLNLSGDIQYAKGASQANSAGYVDADGNKVYDSTD 1480
 Db 1602 TVETEDCKDEGNGLVAKAVYDAVNNKAGWVKT--GANGNDOPAVY--ASGVNTE---- 1655
 QY 1481 NKYYQAKNDGTVDKTEVAKDKLVAQAPDGL--AOMNVKSYINKEQVNDARKKG-- 1536
 Db 1656 -----ADGNCT--TAEVTK-----ANDGSITVYKYNK-----VADGKLKLDKI 1692
 QY 1537 INEDNAVVGLEKRAASDN-----KTKNAVTVGDLNAVAGTPTLPAGDGT----- 1582
 Db 1693 VADTVTVYVADGKYTATNNKXGKFKDASGLACLNKLSAT--ATACKESTGEVDANSA 1750
 QY 1583 -TAKLGEETLLIKGGQTDINKLTDNNIGVYAGDFTVKLAKDLTLMNSV-----NAG-- 1634
 Db 1751 GQEVKAGDKVTFKAG-----DNLKIKQSKKDFYSLKELKLDLSEFKDANGTG 1801
 QY 1635 --GKIDDKGVSPVDSGQA-----KANTVYLSANGLDGKVIYSNVGRGT 1679
 Db 1802 SESTKTKDKGLTTPANGAGAGANTNTISYTKDGISAGNKAVTNVVGSLK 1853

RESULT 14

US-08-685-467-4
 ; Sequence 4, Application US/08685467
 ; Patent No. 6060059
 ; GENERAL INFORMATION:
 ; APPLICANT: St. Geme III, Joseph W.
 ; APPLICANT: Batenkamp, Stephen J.
 ; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 ; STREET: Four Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/685,467
 ; FILING DATE: 22-JUL-1996
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/409,995
 ; FILING DATE: 24-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Silva, Robin M.
 ; REGISTRATION NUMBER: 38,304
 ; REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEFAX: (415) 398-3249
 ; TELEX: 910 277299
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1912 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; US-08-685-467-4

Query Match 7.9%; Score 845; DB 3; Length 1912;
 Best Local Similarity 22.2%; Pred. No. 4,4e-41;
 Matches 478; Conservative 254; Mismatches 648; Indels 772; Gaps 96;

QY 1 MHNIYKVFENKATGTFMVAEYAKSHSTGGSCATGOVGS-----VRLSEFARIALAVL 55
 Db 1 MKKIFVIMNVMTQGVVWVSELTRH-----TKRLNRKDPVLALLRLATYQANF 51
 QY 56 VIGATLNGSAVAGIGISEADGGKGANRDKSAIGDIAQ-----ALGSOSIAGDKI 110
 Db 52 -----DEDELDPPVRAVAPLSPHSDREGTEKEV 81
 QY 111 VHNNSNNANIGAKASGONESIAGDVLASGHASISAGSDDLTKREYQOISELLPIIRG 170
 Db 82 TERNMNGIIFDKNG-----VLKAGATLLKAG--DNLKXKQXTDE----- 118
 QY 171 QKALNDIYQADPNIQKRRTHAQGHASTAVAGMSYAGHFSNAGTRATAGTSLAVG 230
 Db 119 -----XTNASSF-----TYSLKLD 132
 QY 231 LT-ATKAASSIAVGSNMAOIGFPAATAVG-----GSTOVNLNRGIALGFSQVLOKND 283
 Db 133 LLDLITVATREKLSFGANGKQVDTSDANGKLAKTNGVHLN-----GLDSTLDAVTN 187
 QY 284 VNAANVRAVAPDDNCPIDNRYKATFRNGATDYFSIG--NSNGDSIRKLIINAGSADTD 342
 Db 188 TGVLSSSSEFPND--VEKTRATVYK-----DVLNMGWNIGK-----AKTA 225
 QY 343 AVNVAQLKEAVRL--ANROITFKGDSNNRVEKGLKTLITIGAOTSALT--DHNIGVQ 399
 Db 226 GGNV-----BSVDLVSAVYNNVEFITGDKNT-----LDVLTAKENKTTVKTPTSVYK 276
 QY 400 NGDGLVQVLAETLTSLKMYTENTLANKRYTVGRKRLTDTKIGFTDMMNGIDESKPYLDK 459
 Db 277 EKGK-----KLTGKKNNDTKNT--SATADN--TDEGNGLVAKAVID- 317
 QY 460 DTGIIHAGGQITKLAGVVDDAIYQGLKKVNOYAESLQTFYKVKVKN--D 513
 Db 318 --AVNRAGWVVKTTANGONGDEATVA--SGTNVPEESDGTASVTKTKTNGITVYKD 373
 QY 514 A-----NDSKITTVGKNKPKDGOYNTL-----KLKGENGVVVT- 548
 Db 374 AKVGDGLKFPDSKKIYADTALTLYTGKVAEIAKEDDKKLVNAGDLVYALNLSMKAKA 433
 QY 549 --ETNGVTTFGLNQNGNLVGNSTLNNDGLSVK--NTNSNKQIQVGADITFTDISNS 602
 Db 434 EADTDALE--GISKQDEVKAGEVTFKAGKRLVKODGANFTYSLDALTGLTSLTIGT 492
 QY 603 KPGAGIENTRITRDIQGF--ANNTSLPANRRLPTGINGAGKRLTVQASINATNG 660
 Db 493 ING--GNDKATVINKDGLITTPAGNGGTTGTNTISYTKDGIRKAKRAITVNASGLR-AYD- 550
 QY 661 GOLDFMNRIST-----ANTE--KSGSAATIKDLNLSQVPLTPA 697
 Db 551 ANEDVLNNSATDILNRHVEDAYKGLNLNENKANKQPLVDTSTAATVGDRLKLGWVSTKN 610
 QY 698 G-DTGPVNTKIGELIKVKGKTTADDLTKNNIGVAVDSTNSLYVLAKTISDL----- 751
 Db 611 GTFEESNOVQADDEVLTGAGATVTSKSENG-----KHRTIVSAVETADCGLEND 662
 QY 752 -DAVNTTILTA--SDKYVVDGSNNTFAKIQNGDLTFESKONTGATPA----- 793
 Db 663 GDTIKLKVNDQNDNDVLTVGNNGTAATVKG--FEYVTKGATADRGKTYVDATANDAD 719
 QY 794 -----TNSKITGVDLKFTDN--NGIALDGTYYITRKDYFAK 829
 Db 720 KKVATVADVATFAINSATFVKTELTLSIDEDNPIDGKNDALKAGDTLFFKAGKMLKYK 779
 QY 830 QDG--SLDKSKPYLDK-----DKLAVG-----EVEITTINGIN- 859
 Db 780 RDKRNTFFDLAKNLEVETAKVSPDITLITGNTPTGCTAPKVVITSTADLNAKETAADA 839
 QY 860 AGGKAT--TGLSNTLD--ATNATTHGHTQGIYDSTDKTPAASIGDVLNFGNLLKNNGA 916
 Db 840 SGSKNVYLIKGIATLTPSAGAKSSHVDL--NVDATKKSNAASIEDVLRAGMHIQNGNN 897
 QY 917 KDFVSTYDVTDFI----- 929

```

Db 898 VYVATYDVTNFTDSTGTTVTYQKADGKADYKIGAKTSVIRKDHNGKLTFTGDKLDA 957
      :|:|:|:|:|:|
QY 930 -----
Db 958 NNGATVSEDDGKDTGLTAKTVIDAANKSCMRVTEGATGATAVNAGNAETVTS 1017
QY 930 -----NNGATVSEDDGKDTGLTAKTVIDAANKSCMRVTEGATGATAVNAGNAETVTS 1017
      :|:|:|:|:|:|
Db 1018 TSVNKNKNGATTATATYKSKDGNINVKYDVAVGDKL-----GDKKIYADTTITLVGG 1071
QY 983 K-----GDKAIFSVNSGDKALINAKDIDNLNLTAGEIR-----NTKGTADTALQ 1029
      :|:|:|:|:|:|
Db 1072 KVSYPAGANSVN-----NKKLVNAEGLATLNNLSWTAKADKADYDESEGETDQEVKA 1125
QY 1030 -----TFQV-----KKVEN-----
      :|:|:|:|:|:|
Db 1126 GKRVTFFKAGNKLKVKQSEKDTTYSLODPTLGLTSLTGLGTANGRNDGTVINKGLTITL 1185
QY 1040 -----GDDNDADPTITVQD-----
      :|:|:|:|:|:|
Db 1186 ANGAAGATDASNGNTISVTKDQISAGNKEITVNSALKTYYDQNTQADETDQKPHAAVK 1245
QY 1059 QVNTLKLKAGKNGLDIQTNDG-----TVTFGINTQSGKLAGNNTILNNG-----LSIKNTAG 1111
      :|:|:|:|:|:|
Db 1246 NANEVEFGKNGATVSAKTIDNNGKHTVTDV-----AAKVGGLKEDTDGKIKLVNDTIDG 1302
QY 1112 NFOIVGADGVKFAKVNNGVGA-----GIDGTRITRDEIGFAGTNGS-----LDKSKP 1161
      :|:|:|:|:|:|
Db 1303 NMLLV-----DATGASVANGEFMAVTTDATTAOGTANERKGVYVGSNGATATEIDK----- 1357
QY 1162 HUSKDGINAGKKTITNIOGELIAQNSNDAYTGKITYDKTELENKISSTAKTAQNSLHEF 1221
      :|:|:|:|:|:|
Db 1358 -----KKVATV-----GDVAKAINDAAFEVKV-----ENDDSATI-----DD 1389
QY 1222 SVADQGNFTVSNYSSTDSKTSIDVTF-----AGENGITTKYKGV-----VRGIDQTKLT 1277
      :|:|:|:|:|:|
Db 1390 SPTDGGAN-----DALKADPTLTLAKGKMLKVRKROKNTFPALANDLSVKSATV 1438
QY 1278 TPRLTGVNNGK-----GIYIDSONGONT-----ITGLSNTLANVTNKGSRVTEQ 1324
      :|:|:|:|:|:|
Db 1439 SPSLSLGTNGKVNITSDTKGTFPAKDSKTGDDANIHLNGLASTITLDTLLNSGAT-TMLG 1497
QY 1325 GKIIKDEKTRASTIVLVSAGFNLO-----NGEAVDFVSTYDVTNFPADGNATTAK 1376
      :|:|:|:|:|:|
Db 1498 GNGITDNEKKKAAVSKVDYLNAGMNVGVKPPASANNQVENIDFVATYDVFVSGDKDPTS 1557
QY 1377 VTYD-----DTSKTSKYVYDVNVDTTIEVADKLGVTTLTSTG-----TGAN-KFPLSNQA 1429
      :|:|:|:|:|:|
Db 1558 VYVESKDNKGRTEV-----KIGAKTSVIRKDHNGKLTFTGKELKDKANNNGV 1601
QY 1430 T-----GDALVKASDIVAHLNLTSLGDIQTAKAGASQANSAGYVADGNKVIYDSTD 1480
      :|:|:|:|:|:|
Db 1602 TYTEEDGDEKGLVTAKAVIDAVKAKMRVKT-----GANGQDDPFTV-----ASGTNTVF----- 1655
QY 1481 NKTYAKADGVTYDKTEKVAKDLVAAQCTPPGTL-AQMNKVSINKEOVDANKKQG--- 1536
      :|:|:|:|:|:|
Db 1656 -----ADNGT-----TAEVTK-----ANDGSIIVKYK-----VAAGLKLIDGDKI 1692
QY 1537 INEDNAFVKGLEKASDN-----KTKNAAYVYVDLNAVAQPLTFPAGDTGT----- 1592
      :|:|:|:|:|:|
Db 1693 VADTIVLVADGKVTAPNNKGKKTXXDASGLAGCLNKLSKT-----ATAGEGGEVDPANSA 1750
QY 1583 -TAKKLGELTITKGGOTDNTKLTNNINIGVAGTGFVKLAKDLTNLNSV-----NAG-- 1634
      :|:|:|:|:|:|
Db 1751 GQEVKAGKVTFFKAG-----DNLKIKOSXKDFYTSIKKELKDLTSLVEFKANGSTG 1801
QY 1635 -GRTIDKGVSVFVSSQA-----KANTPVLASNGLDLGGKVISNNGKGT 1679
      :|:|:|:|:|:|
Db 1802 SESTKITKDGTLTPANGAGAAGANTANTISVTKDQISAGNKAATVNVVSGLK 1853

```

RESULT 15

```

us-09-268-347-32
: Sequence 32: Application US/09268347
: Patent No. 635182
: GENERAL INFORMATION:
: APPLICANT: Loosmore, Sheena M.
: TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
: FILE REFERENCE: 1038-860
: CURRENT APPLICATION NUMBER: US/09/268,347
: CURRENT FILING DATE: 1999-03-16
: NUMBER OF SEQ ID NOS: 54
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 32
: LENGTH: 1094
: TYPE: PRT
: ORGANISM: Haemophilus influenzae
: US-09-268-347-32

Query Match
Best Local Similarity 25.7%; Score 693; DB 4; Length 1094;
Matches 310; Conservative 159; Mismatches 464; Indels 274; Gaps 53;

QY 1076 NKDGTVFEGINTQSGKLAGNNTILNNGLS--IKNTAGNEQIOVGADGVKFAKVNNGV 1133
      :|:|:|:|:|:|
Db 2 NKIFNVIMNVMTQWAVVSELTRAHTKRASATVATVATOLATASAEANSSASVTSRLNV 61
QY 1134 AGIDGTRITRDEIGFAGTNGSLDKSPHLSKDGINAGKKTITNIOGELIAQNSNDAYT 1192
      :|:|:|:|:|:|
Db 62 YG-DTNKFN-----AANNSIADLNK-----QNDGVDHGLNLTENKANKKLVDNDAAT 111
QY 1193 GKITVDL-----KTELENKISSTAKTAQNSLHEFVADQGNFTVSNPS----- 1238
      :|:|:|:|:|:|
Db 112 VGDRLKLGWVSTKNGKENKSOQVQADLVLFKSGKGVQVSTSENGHATFALAD 171
QY 1239 -SVYFSTSDVITFPAGENGITTKYKGVAVGVGDQTKGLTTPILV-----GNNKKGIV 1292
      :|:|:|:|:|:|
Db 172 LDMRTAVSDPTLTIGGS-----TTSATTPRVNVTSPASGLNFAKAT 215
QY 1293 IDSQNGONTI--TGLSNTLANVTNDKSVRTEQKLIKEDTTRASTIVLVSAGFNQ 1350
      :|:|:|:|:|:|
Db 216 -GANGGTTVHLNINASTLDDTLNLTNGVSKLDONGITADE-KKRAASQVDVLSNGMNR 272
QY 1351 G-----NGEAVDFVSTYDVTNFPADGNATTAKVYDTSKTSKYVYDVNVDTTIEVKDK 1404
      :|:|:|:|:|:|
Db 273 GVKGTATTSNDVDFVRYDVEFLSGSEETLVTVDSSENGSKTKYKIGAKTSVIREKDG 332
QY 1405 KLGKTTTLTSTGTC--AKKPLSNQA-----TGDLVYASDVLVALNLTSLGDIQTAKA 1457
      :|:|:|:|:|:|
Db 333 KL-----FTGANKDTNQVASNNAADPTDEKGLVTAETVINAIVNKRAGWRIKTTGAN 384
QY 1458 SQA-----NSSAGYVADGN-----KVIYDSTDN-----KYYQAKNDG-----TVDKT 1495
      :|:|:|:|:|:|
Db 385 NQAGQFETVAGSINVPFADNGNTTAVYTGATNGITVYKFAKVGDDKLKGNQKITADTT 444
QY 1496 K-EYAKDKLVAAQCTPDGTIAQNVKSVINKEOVDANKKQGINEDNAFVKGLEKASDN 1554
      :|:|:|:|:|:|
Db 445 ALTVTGKVTAT-----PDAT-----NGKKLVNASGLAALNK-----LSMTAKAEAD- 486
QY 1555 KTKNAATVGDINAVAQPLTFPAGDTGTAKKLGELTITKGGOTDNTKLTNNINIGVAA 1614
      :|:|:|:|:|:|
Db 487 -TANG-----GELD-----GTADKEVEKAGETVTFKAG-----NLAKYKDG 522
QY 1615 DGFYVLKADLTNINSV-----NAGCTKIDKGVFVSSQA--KANTPVLASNGLDL 1666
      :|:|:|:|:|:|
Db 523 ANFTYSLQDALTLGLTSLTGTGNNGAKTEINKDGLITTPANGAGANNANTISVTDGISA 582
QY 1667 GKVYISVNGKGTDTDAANYQOLNEVRNLL-----GLNAGNDNADNQOVNIAD-- 1715
      :|:|:|:|:|:|
Db 583 GQGSVKNVVSGLKFFGANDPPLTSSADNLTQYDDAYAGLTNLDKAGADKOTLVYADNT 642
QY 1716 -----IKKDPNG-----SSSNFTYIKAGTV--LGKGNNDTEKLANTGQVQV 1757
      :|:|:|:|:|:|
Db 643 AATVGLRLGLGWISADKTTGELDKKEYNAQVRNANEFKRSNGINVSGLTYNGRREITF 702

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:35:52 ; Search time 51.53 seconds
(without alignments)
3956.949 Million cell updates/sec

Title: US-09-813-214A-9

Perfect score: 10708

Sequence: 1 NMHITKVFENKATGTFMVA.....NCSADTQGHVGAAGACGPHF 2122

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|--------------------|
| 1 | 1246.5 | 11.6 | 2059 | 2 D82671 | surface protein XF |
| 2 | 986 | 9.2 | 1588 | 2 A86036 | probable adhesin Z |
| 3 | 986 | 9.2 | 1588 | 2 H91188 | probable adhesin E |
| 4 | 728 | 6.3 | 1190 | 2 A82615 | surface protein XF |
| 5 | 670 | 6.3 | 1107 | 2 AC0976 | probable autotrans |
| 6 | 669 | 6.2 | 3705 | 2 AD0123 | probable autotrans |
| 7 | 651.5 | 6.1 | 2340 | 2 B71704 | cell surface anti |
| 8 | 645.5 | 6.0 | 5291 | 2 B85547 | hypothetical prote |
| 9 | 634 | 5.9 | 5188 | 2 A41477 | probable RTX fami |
| 10 | 630.5 | 5.9 | 2249 | 2 B89921 | 190K surface anti |
| 11 | 622 | 5.8 | 6713 | 2 B89921 | hypothetical prote |
| 12 | 598 | 5.6 | 3013 | 2 AB0480 | probable invas |
| 13 | 590.5 | 5.5 | 2271 | 2 F90073 | hypothetical prote |
| 14 | 568.5 | 5.3 | 2554 | 2 AB3528 | extracellular seri |
| 15 | 550.5 | 5.1 | 2021 | 2 A97859 | 190-KDa cell surfa |
| 16 | 549.5 | 5.1 | 2535 | 2 AC0304 | probable hemolysin |
| 17 | 548 | 5.1 | 4919 | 2 T31105 | hypothetical prote |
| 18 | 546.5 | 5.1 | 2020 | 2 C48399 | ABC-type transport |
| 19 | 541 | 5.1 | 2660 | 2 E85822 | probable invas |
| 20 | 539.5 | 5.0 | 2703 | 2 H81193 | hemagglutinin/hemo |
| 21 | 536.5 | 5.0 | 4152 | 2 T31102 | filamentous hemag |
| 22 | 534 | 5.0 | 3624 | 2 AD0835 | large repetitive p |
| 23 | 533 | 4.9 | 658 | 2 AH0110 | probable surface p |
| 24 | 530 | 4.9 | 3890 | 2 C89921 | hypothetical prote |
| 25 | 529.5 | 4.9 | 1910 | 2 AF0394 | probable adhesin h |
| 26 | 527.5 | 4.9 | 5627 | 2 C83339 | hypothetical prote |
| 27 | 521 | 4.9 | 3295 | 2 AE0074 | probable adhesin Y |
| 28 | 519.5 | 4.9 | 2468 | 2 A83412 | hypothetical prote |
| 29 | 517.5 | 4.8 | 4936 | 2 AH2515 | hypothetical prote |

| | | | | | |
|----|-------|-----|------|----------|--------------------|
| 30 | 517 | 4.8 | 2481 | 2 D90011 | FmB protein [limp |
| 31 | 517 | 4.8 | 3029 | 2 S76109 | hypothetical prote |
| 32 | 513 | 4.8 | 2514 | 2 F81045 | hemagglutinin/hemo |
| 33 | 508 | 4.7 | 1536 | 2 A43855 | high-molecular-wt |
| 34 | 508 | 4.7 | 2383 | 2 D64962 | probable membrane |
| 35 | 500.5 | 4.7 | 1651 | 2 JC1340 | outer membrane pro |
| 36 | 489.5 | 4.6 | 1477 | 2 B43855 | high-molecular-wt |
| 37 | 487 | 4.5 | 2479 | 2 F87386 | conserved hypobol |
| 38 | 483 | 4.5 | 1999 | 2 AB2018 | hypothetical prote |
| 39 | 480 | 4.5 | 1643 | 2 D71630 | outer membrane pro |
| 40 | 478.5 | 4.5 | 3591 | 2 S21010 | filamentous hemagg |
| 41 | 477.5 | 4.5 | 2055 | 2 T31110 | extracellular matr |
| 42 | 477.5 | 4.5 | 3535 | 2 E83641 | probable hemagglut |
| 43 | 476 | 4.4 | 2893 | 2 A64556 | toxin-like outer m |
| 44 | 474.5 | 4.4 | 1577 | 2 A35140 | hemolysin A precu |
| 45 | 467 | 4.4 | 1487 | 2 AG2560 | hypothetical prote |

ALIGNMENTS

RESULT 1
D82671
surface protein XF1529 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence: revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: D82671
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: D82671
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2059 <SIR>
A:Cross-references: GB:AE003982; GB:AE003849; NID:g9106554; PIDN:AAF4338.1; GSPDB:GN
A:Experimental source: strain 9a5c
R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briñones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrero, D.M.; Carre
as-Neto, E.; Docena, C.; El-Dorri, H.; Facinca, A.P.; Ferreira, A.J.S.
Submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Klatjima, J.P.; Krieger, J.E.; Kurama, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1529

Query Match 11.6%; Score 1246.5; DB 2; Length 2059;
Best local Similarity 24.3%; Pred. No. 1.6e-40;
Matches 578; Conservative 344; Mismatches 780; Indels 681; Gaps 108;

62 NGSAVYAGIGISEADG-----KGGANRAGDKSTAI-----GDIQA 97
36 NAOYVINDGM---DGGCQRIYDNGSASGVRYVATQCSDEPTVTGYARFGSGIAAD 92
98 LG-SOSIAIDNKNIVHNSNNANIGAKASGNEIAIG-GDVL-----ASGHASTAIGSDDL 151
93 QGASRNLTIGSLYV---NSGOVGVNDVNTKTYIRMSGVITIMNTVAGNATAIGS--- 145
152 YLKREYVOQISELLPIIRGKALNDIYOLADPTWLOKRRYRHAQGHASTAVGA--MYANG 209
146 -----AASSADALKAQLATKAG 164
210 HFSNAFTAFAGTYSLAVGTATATA-KAASIAVGSNAQAIGFAATVAVGSGTQVNLNG 268

```

Db 165 ARAIALGAKASADVPFVALGSGATAGTASSIALGLMSAVN-GAVAVGGALVYVPBG 223
Qy 269 -IALGGSOYLQKNDVNAANVAIYAPDNOPIDNRYKATFKNGATFVPSIGNSGNDSI 327
Db 224 AVAALGNSVA-----STGKLSGSDP-KYKTTSTDSAAKMSLTAAV-SIGDVSSTNKK 275
Qy 328 RRIIVNGASADTDVAVNAOLKEAVRLANR-QITEFKGDSDNRYVE-----KGLGKLT 379
Db 276 TROLSLAGTSMTDVAVNAOLKVDEIASRGNLTAASGANSNAVPGSSVDLKNITDKNL 335
Qy 380 TIT-----GGAQTSALTDHINIGVONGD-----GLKVOLET-----LTLKWTTE 421
Db 336 TITKALGSDNVOFNKNDKYVTTLAVGDALINTDIALGFDVSLSTGLAITDGPVAVS 395
Qy 422 NLTANEKV-----TVGKTRLTBKIG-----FTNDMNGIDES----- 453
Db 396 GIDAGSKVISHVAAVSEI-STDVAVNGSOLNAVQVQASOPVTFGNGGAVKRSIGOSV 453
Qy 454 -----KPLDKOTG-IHA-----GGOKITKLTAQVVD 481
Db 454 VISGESSTAGTYSGNLKSVDPAARHILQADSPKFGNVYINNGKISCTAGTEFD 513
Qy 482 AATYGLKKNVQTAESALQTFYKVKYKNGNDANDSKITTYGKNKPDGTQVNTLKLGE 541
Db 514 AVNFSQLKSISTAYD--OGWTLTASGNG--SKVASG-----TYDLKNT 554
Qy 542 NGDVYTFENG--TYTFGLNQN--NGLVNGSTINNGLSKKNNSKQIQVQADGT 595
Db 555 DG-NLITISGSDNDVAVNFSLKDFGMSGTIVVNDGKVR--GSDVALGTTGLT 608
Qy 596 FTDISNSKPGAGIENTRTTRDGIFFANNTGLDANKPLTPTGLINAGKELTNVOSAI- 654
Db 609 ITD-----GPAVTHASGIDAGSKVISHVAAV 635
Qy 655 -----NPAINGGOLDENRULSTANTEKSGSAATIKLYNSQV--PLTFADDTGPNVK 706
Db 636 SETSTOAVNGSQL-----NAVQVQASOPVTFENGGA-VKR 670
Qy 707 KIGELIKVGGKTTADDLKNIGVAVSDTNSITVYKLAKT----- 747
Db 671 SLDGSVVISGESSTAGMYSGNLSKYVDEAGRIHQLADSPKFGNVYINNGKISGVTA 730
Qy 748 -LSDDAVNTKL--TASDK--VTVDGNNNTAKL-----QNGDLTFKSKONTGA 790
Db 731 GFETDAVNFSQLKSISTAYDQWTLTASGNSKVASGCTVDLKNITDNLITISGSDN 790
Qy 791 TPAITN-SKITGVGKFTDNNGLALDGTITRKD-KVGFPAQDGSGLSKFYLDKDKL 847
Db 791 DVVFNLSKDFKVDG-----TSGTIVVNDGVKVG-----SDVALGTTGLT 831
Qy 848 VGE-VEITTINGINAGKAITG-----LSNTLPDATNATGHTVGLGIVDSTKTRAASIG 901
Db 832 ITDGPVAVTHASGIDAGSKVISHVAAVSEISTDVAVNG--SOLNAVQVQASOPVTFG 886
Qy 902 DVNLAFNLSKNNNDKADPVSITVDVFINGNATTAFTVYDGKASKVAAYDVNTGTTIHLT 961
Db 887 ---NEGAVKRSIGOSV-----VISGESSTAG-TYSGNLSKYVDEAAG--RIHLQ 930
Qy 962 GADGNKNQIGVKTTLTKTDAGDKRAINFVSNGDDKALINAKDLADNLTLAGELRN-T 1020
Db 931 LADSPF-FG-----NAVJING-----GRISGVT 952
Qy 1021 KGTADP-ALOTFOVKKVEKNGDDDDNDADITVYGAKDAKTQVNTLKLKNGLDIOTKNDG 1079
Db 953 AGTEFDVAVNFSQLKSISTAYDQWTLTASGNSKVASG--G 993
Qy 1080 TYTFGINTOSGLKAGNNTLNNGLSIKNTAGNEOIQVQADG--VKFAKVVNNGVAGAGID 1137
Db 994 TV-----DLKNTDGNLITISGSDNDVAVNFSLKDFGMSGTIVVNDGKVR--GSDVALGTTGLT 1030
Qy 1138 GTTRITRDEIGRA-----GTNSGLSKRPHLSKDGINAGGKITTINQSEIATONSNDAYT 1192

```


RESULT 2
 A86036
 Probable adhesin 25029 [imported] - Escherichia coli (strain 0157:H7, substrain EDL933)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: A86036
 R:Perma, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Iller, L.; Grobleck, E.J.; Davis, N.W.; Lim, A.; Dimmlanta, E.; Potamocistis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.
 A:Reference number: A85480; NCID:21074935; PMID:11206551
 A:Accession: A86036
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1588 <STO>
 A:Cross-references: GB:AE005174; NID:912518349; PIDN:AAG58749.1; GSPDB:GNO0145; UMGCP:Z50
 C:Experimental source: strain 0157:H7, substrain EDL933
 A:Gene: Z5029

Query Match 9.2%; Score 986; DB 2; Length 1588;
 Best Local Similarity 22.9%; Pred. No. 1.3e-30;
 Matches 510; Conservative 273; Mismatches 696; Indels 752; Gaps 95;

QY 1 MNHIVYFNKATGTGMAVAEYAKSHSTGGSCATGQVGSVRLSPARIALAVLVIGAT 60
 1 MNKIFKYMNPATGNTVYTSKRSKRSK-----LLISALVAGM 45
 61 LNSAYAGIGISEADGKG---GANARGDKSIAIGDIAQALGSSIAIGNKIYHNSNN 117
 46 L--SSFALANAGNDNGQVDYSGSAGDGMVAIGKAKA-----NTFMNTSGSS 93
 QY 118 ANIG--AKASGNEISAIAGDVLASGHASIAIGSDLYIKETQYISELLPTIRGQALN 175
 94 TAVGYDAIEQYSAISGSKTHAIGASMAF-----VSAISE----- 131
 176 DIYOLADTNLOKRYRTHAOGHASTAVGAMSYAKGHFSNACTRTATAGTYSLAVGLTATA 235
 132 -----GDRSIALGASSYSLGQYSMALGRYSKALGKLSIAMGDSKA 172
 QY 236 KAASIVAGSMAOAIIGFAATVAGSGTOVNLKRGIALFGSOVLQKNDVNAANRATAPD 295
 173 EGANAIALGNATKATETMSIALGDTANASRAYSMALG-ASSVASEENAI-AIGETPAAE 230
 QY 296 DNOPIDNRKATFKNG-----ATDVFISGNSGNDISIRKTIINAGSADTDVAVNA 347
 231 NATIIGNAKAKKGTNSMAMGFGSLADKYNTIALGNSOLADNAIALGQG-----NKA 283
 QY 348 QLKRAVRLANROITFKGDDSNRRVEKGLKTLITTCAGQTSALTNDHNIQVONGDGLKVQ 407
 284 DGYDAIALGN-----GSQS-----RGLN--TALGTASNATGDKSLAGSNS----- 324
 QY 408 LAEFLTSIKAVTTNLTANEKRYVYKTRLTGKTIGTNDMNGIDESPYLDKOTG--IHA 465
 325 -ANGINSAIALGADSIADIDNTVSGNSSL--KRIIVNVKGAIKISDY--DAINGSOLYA 379
 QY 466 GGGKITRTKTAG--VYDDDAATYGGLLKRVNQTAESALOTFYKVKDKKNGNAN--DSKI 519
 380 ISDSVAKRLGGGAAYVDDGTV-----TAPTYNLKNGSKNNVGAALAVDENT 427
 QY 520 I-----TVGKNNKPDGTQVNTLKLKENGVDVTTETNGTVFGLNQNNGLTVGNSTLNDG 575
 428 LQMPQTKKYSAAHGTSSPTASV-----ITVDAGTIS----- 460
 QY 576 LSVKNTSNKQIOYGACDITFTDLSNKPAGAGIENTRITRFDGIGFANNGSLDANKPRL 635
 461 ASSSDAVNGSOLKATNDVE-----ANTANIANININIANIANTNANTNTNI 507
 QY 636 TPTGINAGKELTIVQSAINFATNGGOLDFMNRISTANTKESGSAATIKDLYNLSQVPLT 695

DB 508 T-----NLDSVGLDQ--ADALLMNETKFAFSA----- 534
 QY 696 FAGDGPNNYFKLGLGELLKYGCKTTADITFKNNIGVADSDNSLTVKLAKTSLDLAVN 755
 535 F--HGDPTSK--ITNNKDA-----DLT-----ADSTD-AVNGSOLKTTTNAVATN 574
 QY 756 TKTLTSDKVTYVDSGNNTKLQNGDLTFESKQNTGATPAPNSITIGVDGLFDNNGI--A 813
 575 TINI-----ANNISNIAKNTNTNISLITEVT-----NLGEDALKWKDNGVFA 618
 DB 814 LQGTYYITRKDYGFAKQDSDSLDK-SKPYLDKDKLKYGEVETTINGINAGKATIGLSMTL 872
 619 AHQTE--TISKITNVR-DCDLTTGSDAVNGSOLKTTNDAVAFNTNTIATNT--TINSLT 674
 QY 873 TDATN-----ATTGHVQLGYVSDTIDKTRASIGDVLNAG--FNLKNG 914
 DB 675 EYVTNIGEDALKWKDNGVFAAHGNMTASKITNILDGVTATSSPAINGSOLYLSSN- 733
 QY 915 DADPFSTVDYDFINGNMTAKVYDCKASKAVADVNDGTHILTGADGKKNGIGVKT 974
 734 ---IATY---EGGNAS-----VNDGVFTGPTYKIGETNNY-- 765
 QY 975 TLLTKTDAGD--KAIN--FSVNSGDDKALINAKDIADMLNTLAGEIRNTKGTADTALQT 1030
 766 -----GDALAIANSFSTSLD--ALL-----WDATAGKFSKHT----- 799
 QY 1031 FOYKVKVKGND-----DDNDADITYGKADKTNQVNTLKKRN-----GLDIQTKD 1078
 800 -----NGDASVITVDVADGEI-----SDSSSDAVNGSOLHGVSSVVDALGGAEVND 847
 QY 1079 GTVTFPIINTQSGAKAGNNTLNNNG--LSIKMTAGDEQIQVAGDGVKFAKVVANNVYAGI 1136
 848 GTTATPTIYA-----NADYDVGDALNAIDITLDDALMDAD----- 885
 DB 1137 DGTTRITRDEIGFAGTNGSIDKSPHLSKDGINAGKKTITNIOGIEIAONSNDAYTGKI 1196
 886 -----AGENGAF--SAAH-----GMDKRTASVITVANAGISAASDAINGSQL 926
 QY 1197 YDKTELENKISSTAKTAQN--SLHESVADBOGNFTVSNPYSSYDTSKSDVITFRG 1253
 927 YTNKXIADALGDAVNAVDGTTATPTIYIANAVERN-----VGDMALDALD 972
 QY 1254 ENGITTKVKKGVVRVLDQTKGLTTPKLVNNGNGIYIDSONGNTITGLSNLTANT 1313
 973 DNALLW-----DETA-----NGGAGVNAHSDGKAST-----ITNVA 1004
 DB 1314 NDKGSVTEEGKIIIDENKTRASIVDVLSAGFNLOGGEAVDFVSTDFVNFADGNAT 1373
 1005 N--GST-----SEDSST-----DYNNGSQLANT 1024
 QY 1374 TAVTYDDTSKTSKYVDV--NVDDTTIE-----VKDKLGVKTTTLTSTGTGANK 1423
 1025 NMI-----EONQIINOLAGNTDATTYIOENGAGIYVTRNDGGLAFNDSAGGATAI 1079
 QY 1424 ALSNOATGDAV-----KASDIYAHNLTLSDGTQT--AKGASQANSSAGYDADONKVI 1475
 1080 GYNSVAKGDSVAALIGOGSYSDVDTGIALGSSVSYSVIKAGSDYS--TTENGVLVG 1134
 QY 1476 YDSTDNKKYQAKNDGTIVDKTRKAEVAKDLVAQAOTPGTLAOMNVKSVINKEOVNDANKRO 1535
 1135 YDTTDEELGALSIG-----DQCKYKO 1156
 QY 1536 GINEDNAFYKGLKKAASDNKTKNAATVGDV--NAVAOTPLTF--AGDTGTAKKLGE 1589
 1157 IIN-----VADGEADDAVTVROLONAIAGAVATTPTKYPHANSTEDSLAVGT 1204
 QY 1590 TLTIKGQDPTNKLPTDNNITGVVAGT-----DGFYVLAKLQTLTNSNAGGTRKIDKG 1642
 1205 DSLAMGAKTIVN--GKGKIGICGAYVANALNGIAGISNAOVIHNSAIGNSGTTTRG 1262
 QY 1643 VSF-----VDSGQAKAMPVYLSANGLDLGGKVISNGKGTKDDAANVQOLNEVRNLL 1696
 1263 AGNTNYATNMDAPONSVGERSVGSAD-----GQRQITNVNAGSADTDVAVVGLD----- 1311

Oy 1697 GIGNAGNDNADNOVNIADIKKDPNSGSSNRRTYKAGTVLGKGNNDTEKLTAGVOVG 1756
 Db 1312 -----KVDAQVSQNTQSTITN-----LDNRVTNIDSRVT--NIENG 1345
 Oy 1757 VDKDNANGDLSNVVVKTKQDKSGKALLATYNAAGOTNYLTNNPAAIDRINEOGIREFH 1816
 Db 1346 I-----GDIY-----TGSTKYFKTN----- 1361
 Oy 1817 VNDGNOEPVOCGRNCIDSSAGSKHSAIGFOAKADGEAAVAIGRTOGNOGNOIAIDMAQ 1876
 Db 1362 -----TDGVDASAGCKDSVAL-----GSGSIAADN-- 1387
 Oy 1877 ATGDOSIAGTGNVAGKHSAGIADPSTVKADNSYSVGNNOFTDAGOTDFEGVGNNTIV 1936
 Db 1388 -----SVALGTSV-----ATENENTISVGSSTNQ 1411
 Oy 1937 TESNSVALGNSAISAGTHAGTQAK-KSDGTAGCTTTTGAATGYKGFAGOTAVGASVGA 1995
 Db 1412 RRTTVVAGKNA-----TDAAVVAQLKSSSEAGVRYDTRKADGSDID--YSNTILG---GG 1460
 Oy 1996 SGAERIONVAGVSAVSATSTAVNSOLYKATOGIANATN---ELDRHIONENKAMAG 2051
 Db 1461 NGGTRISVNSAG---VNNNDVYNTAOLKOSYQETKQYTDORAWEMDKLSTESKLSGG 1517
 Oy 2052 ISSAMAMASMPQATIPGRSMYTGATIHNGOGAVALGLSLDNGQWFKINGSADTQGH 2111
 Db 1518 IASAMAMGLPQATYPGASMSASIGGTYNGESAVALGVSMVANSRMYWYKLGSTINSQGE 1577
 Oy 2112 VCAVAGAGHF 2122
 Db 1578 YSALAGAGIOW 1588

RESULT 3
 H91188
 probable adhesin ECS4480 [similarity] - Escherichia coli (strain O157:H7, substrain RIMC)
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 02-Nov-2001
 R:Accession: H91188
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yochoyama, K.; Han, C.G.;
 gisawa, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
 A:Reference number: A9629; MUID:21156231; PMID:1158796
 A:Accession: H91188
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1588 <NAV>
 A:Cross-References: GB:BA000007, PIDN:BA837903.1; PID:913363955; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RIMD 0509952
 C:Genetics:
 A:Gene: ECS4480

Query Match 9.2%; Score 986; DB 2; Length 1588;
 Best Local Similarity 22.9%; Pred. No. 1.3e-30;
 Matches 510; Conservative 273; Mismatches 696; Indels 752; Gaps 95;

Oy 1 MHHITKVTNKATGTFMAVAEYAKSHSTGGSCATGQVGSVTLSPARIALAVYIGAT 60
 Db 1 MHHITKVTNKATGTFMAVAEYAKSHSTGGSCATGQVGSVTLSPARIALAVYIGAT 60
 Oy 61 LNSGAVAGISSEADGCGK--GANARQDKSIAIGDIAOALGSOIASIAGDNKIVNSNN 117
 Db 61 LNSGAVAGISSEADGCGK--GANARQDKSIAIGDIAOALGSOIASIAGDNKIVNSNN 117
 Oy 46 L--SSFGALANAGNDGOGVDYSGSAGDGVAIGKAKA-----MFMNTSSSS 93
 Db 46 L--SSFGALANAGNDGOGVDYSGSAGDGVAIGKAKA-----MFMNTSSSS 93
 Oy 118 ANIG--AKASGNESIAIGDVLASGHASIAIGSDLLYKKTVOOISLPLIRGQALN 175
 Db 118 ANIG--AKASGNESIAIGDVLASGHASIAIGSDLLYKKTVOOISLPLIRGQALN 175
 Oy 94 TAVGDAIAEGGYSSAIGSKTHAIGASMAFG-----VSAISE----- 131
 Db 94 TAVGDAIAEGGYSSAIGSKTHAIGASMAFG-----VSAISE----- 131
 Oy 176 DTYQLADTLOQYRRTTHAGHASTAVGAMSTAKGHFSAFGRAIAEGYISLAVGLTATA 235
 Db 176 DTYQLADTLOQYRRTTHAGHASTAVGAMSTAKGHFSAFGRAIAEGYISLAVGLTATA 235

Db 132 -----GDRSIALGASSYSLGQYSMALGRYSKALGLSLAMDSSKA 172
 Oy 236 KAASIVAGSNAOAIGFATAVAGSTOVNLRNGIALGFGSOVLQKNDPNAANVAYAPD 295
 Db 173 EGANAIALGNATKATETLMSTALDITANASKAYSMALG-ASSVASEENAI-ALGAEFEAE 230
 Oy 296 DNOPIIDRRYKATFRNG-----ADVESIGNSGNDISIRKIIIVGAGSADTDAVNA 347
 Db 231 NATAIGNNARAKGNSWAMGFGSLADRVNTIALGNSOALADNAIAIGOG-----NKA 283
 Oy 348 QLEAVVLANRQITFFGDDSNRVEKIGLKTITIGGAQTSALTIDNIGVONGDLKXQ 407
 Db 284 DGYDALGN-----GSSO-----KELN--TIALGTSMAATGDKSLDGSNS----- 324
 Oy 408 LAETILSKAVTTEMLTANKVTVGKTRITTDKIGFTNDMGIDESKPYLDKDTG--IHA 465
 Db 325 -ANGINSVALGADSIADLDITVSGNSL--KRLIVNKNKAIKSDSY-DAINGSLVA 379
 Oy 466 GQOKITKTAG--VDDDAATYQOLKKNVQTAESALQITTVKRVKNDAN--DSKI 519
 Db 380 ISDSVAKRLGGAAVDVDDGTV-----TAPYMLKNGSKNNYGALAVIDENT 427
 Oy 520 I---TYGKNNKBDGQOVNLTKLKGENGVDTETETGVTFGLNQNNGLTVGNSTLNDG 575
 Db 428 LQWDQTKGTSAAHGTSPTASV-----TIDVADGTS----- 460
 Oy 576 LSVKNTNSNKOIGVADGITFTDINSKPGAGIENTRITRDIGFANNTGLDANKPRL 635
 Db 461 ASSKDAVNGSOLKATNDVE-----ANANLANTNTSNLATNTANTATNTNTNI 507
 Oy 636 TPTGINAGKELINVOGSAIPATNGOLDPNRLSTANIEKSGSAATIKDLYLSQVPLT 695
 Db 508 T-----NLTVSVGLQ--ADALLNNEKTKARSA-- 534
 Oy 696 FAGDTGPNVTKLIGELLYKVGKTTADLTNNKNGVADSTDSILVAKLTSLDVAVN 755
 Db 535 ---HGDDTTSK--LTNVSDA-----DLT-----ADSTD--AVNGSOLKTTNDVAVN 574
 Oy 756 TKTLTASDKTVDSGNNTAKLQNDLTFSSKONTGATPATNSKITGVGLKFTDNNGI--A 813
 Db 575 TTNI-----ANNTSNLANTNTNISNLTEVT-----NLGSDALKKWDKNGCVFTA 618
 Oy 814 LDGTTTYTTRKQVRAKDDGSLDK-SKEYLKKLKVGEVITTINGINAGKAITGSLNLT 872
 Db 619 AHGTE--TTSKITNVK--DGLTLTGSDPAVNGSOLKTTNDVAANTNTNIATNT--TNSNLT 674
 Oy 873 TDATN-----ATTGHVTOIGYDSTPTKRAISIGVLIAG--FNLKNG 914
 Db 675 ETVTNLGEDALKKWDKNGCVFTAAHGNNTASKITNILDGTVTATSSDAINSOLYDLSN- 733
 Oy 915 DAKDFVSTYDTVDINGNATTAATKAVYDGAASVAYDVNDGTTIHLTGADGNKNOIGVKT 974
 Db 734 ---IATY---FGGNAS-----VNTDGVFTGPTLYKIGETNYNV-- 765
 Oy 975 TTLTKTAKGD--KAIN--TSVNSGDDKALINKADIADNLTLAGELRNKGTADALQT 1030
 Db 766 -----GDAIAAIISSPSTSLGD--ALL-----WDATRAKFSAKHGT----- 799
 Oy 1031 FOYKVKENGD-----DNDADTTVGDAKTQVNTLTKLKGK--GDIOTFND 1078
 Db 800 ---KDAVITTVADGEI-----SDSSDAVNSQLHGVSSVYVDLGGGAEVND 847
 Oy 1079 GTVTEGINTOSGLKAGNNTLNNNG--LSIKNTAGNEQIOVGADGVFAKVNNGVAGI 1136
 Db 848 GTTAPPTTIA-----NADYDNGDALNIDTTLDDALLMDAD----- 885
 Oy 1137 DGTTRITRDELIGFAGTNGSLDKSKPHLSKOGINAGKKTINIOSGELAONSNDVYTGKI 1196
 Db 886 ---AGENGAF--SAAH--GKDKTASVIINVANGAISASSDAINSOL 926
 Oy 1157 YDKTELENKISSPAKTAON--SLHEFSVADEGNGFTVSNPSSYDTSKTDSDVITPAG 1253
 Db 927 YTTNKYIADALGDAEVNADGTTAPTYTTANAEYN-----VDALDALD 972

```

QY 1254 ENGITTKVKKGVVRVIGIDOTKGLTTPKLVGNNGKGIIVDSONGONTITGLSNTLANVT 1313
Db 973 DALLM-----DETA-----NGSAGAVNASHDKASI-----ITVNA 1004
QY 1314 NKGSGVRTTEOGKIIKDEKTRAAISIVDVLASGFMLQNGEAVDVTYDTPVNFADGNAT 1373
Db 1005 N--GSI-----SEDS-----DAVNSQQLMAT 1024
QY 1374 TAKVYTDTSKTSKVVYDV--NVDDFTLE-----VKDKLVKTTTLTSTGANKF 1423
Db 1025 NMMI-----EQMTQIINLACNTDAFYIOENGAGINVTNTDGLAFNDASAOGVATAI 1079
QY 1424 ALSNATGGAIV-----KASDIVAHLNLTSLGIQ-----AKGSQANSSAGVYDADGNVI 1475
Db 1080 GYNVSAKGDSSVAIGOSYVDVTGIALGSSSVSRVIAKGRDTS-----ITENGVIIG 1134
QY 1476 YDSTDNKKYQAKNDGTVDKTEKAVAKDLVAQAOTPDGTLAQMNVKSVINKEQVNDANKQ 1535
Db 1135 YDTTDELLGALSIG-----DDGKTRQ 1156
QY 1536 GINEDNAFVYKLEKASDKTKNAAVTYGDL---NAVAQPLTF--AGDTGTAKKGE 1589
Db 1157 IIN-----VADSEAHDAVTVYROLQNAIGAVATPTPKYFHANSTEDSLAVGT 1204
QY 1590 TLTIGGGDTNKLTDNNIGVAGT-----DGFYKLANLJTNLSVNAAGTKRIDKG 1642
Db 1205 DSLAMGAKTIV--GDKGIGYGAIVDANALNGIALGSSNAOVHVNSTAIKNGSTTRKG 1262
QY 1643 VSF-----VDSGQAKANTPYLSANGDLGKRVISNNGKTKDTPDAANVOOLNEVRNL 1696
Db 1263 AGNTYATYAMDAAPONSVEGFSVGSAD---GGRITTNVAGSADPDVAIVGGL----- 1311
QY 1697 GLGNAGNDAGNQVNIADIKKDPNSGSSNRTVIKAGTVLGKGNDEKATGCVQYG 1756
Db 1312 -----KVTDAQVSOMTOSITN-----LDNRVTNLDISRYT--NIENG 1345
QY 1757 VDKDGNANGDLSNVWVKTKQKDSKKALLATYNAAGOTNYLTNNPAALDRINEQGRFPH 1816
Db 1346 I-----GDIV-----TTGSKTKFKTN----- 1361
QY 1817 VNDNOEPVVOGRNGIDSSASGKHSVAIGFOAKADEAAVVAIGRTOAGNOSIATGDNQ 1876
Db 1362 -----TDGVDAASAQKDSVAI-----GSGSIADN-- 1387
QY 1877 ATGDGSIATGTGNVAGKHSAGIADPSTYKADNSTSVGNNOFTDATQTDVFGVGNNTV 1936
Db 1388 -----SVALCTGSV-----ATEENTISVGSSTNQ 1411
QY 1937 TESNSVALGSNSAISAGTHAGTQAK--KSDGTAGTTTAAATGTVKGFAGOTAVGANSVA 1995
Db 1412 RRTTNVAAKNA-----TDAVNVAAQLKSSBAGGVRYDTRKDSID--YENITLG--GG 1460
QY 1996 SGAERRIONVAAAGEVASTSDAVNGSOLYKATQIGIANATN---ELDRHIHONENKANAG 2051
Db 1461 NGGTTTISNVSAG---VNNNDVYVNAOLKOSVETKOTYDORVENDKNLSTESTLSGG 1517
QY 2052 ISSAMAMASPPQAIIPERSKVTGCIATHHGCGAVANGLSKLSNDGQWVKINGSADTQCH 2111
Db 1518 ISAMAMATGILPQAYTPGASMASIGGTYNGESAVALGVSVANSANGRMVYKLOGSTNSOGE 1577
QY 2112 VGAAVGAGFHF 2122
Db 1578 YSALALGAGIOW 1588

```

RESULT 4
A82615
surface protein XF1981 [imported] - xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: A82615
R:anonymous, The xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

```

Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen xylella fastidiosa.
A:Reference number: A82615. MIMD:20365717
A:Note: For a complete list of authors see reference number A59328 below
A:Accession: A82615
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1190 <Sim>
A:Cross-references: GB:AE004017; GB:AE003849; MID:99107083; PIDN:AAF484783.1; GSPDB:GN
A:Experimental source: strain 9a5c
R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
as-Vello, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Perito, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Jungelira, M.L.; Kemper, E.L.; Kitejima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeiri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1981

```

Query Match 6.88; Score 728; DB 2; Length 1190;
Best local similarity 23.33; Pred. No. 8.3e-21;
Matches 354; Conservative 191; Mismatches 516; Indels 458; Gaps 59;

```

QY 707 KIGETLKXKGGKTTADDLTKNNIGVADSTNSLTVKLAKTSLDLDVNTKTLFASPKVT 766
Db 27 KIGSLQ--HAWTSAASSK--GTQPRRNNAATAKRSALHRRQLHYLLVLAAT 82
QY 767 VDSGNNFAKLO--NGDLT-----FSKONTGAPATNSKT---IGVDLKTFTDNNGIA 813
Db 63 GYTGKVAQVYVNSDSTENCVEILGSSQTSFHSASMDCKPDTQTEYSLFYDRNLV 142
QY 814 LDGTYITRKVKYKQDSLDKSKPYLDKRLKAVEYEITNNGKRAITGLSNTLT 873
Db 143 LGGSTLYVNEGKLGIVDISGA-----TYSMLRGS--ATMNG--SAGIDSIAGSOGS 191
QY 874 DATNATGTHVTLGLVSDTDKTRASTIG-----DVLNAGFNILKNGDKADKDFSTYDT 925
Db 192 KTDGNTSGATVAAQGL-----RSTAICTIARSOSQDAISIGTGASTGTAALAIAGCAL 244
QY 926 VDFINGNATTAKVYDYDKASKAVAYDVNDGTTIHLTGADGNKKNQIGVKTTLTFTDAKD 985
Db 245 TSIANGIALGASSVTTTGGVALAGGGLAATAASGITVD-----PVTKSTS--- 290
QY 986 KAINFSVNSGDDKALINAKDIADINLTLAGETRNKTGA---DT--ALCTQYKVKYENG 1040
Db 291 -----TLTSMWRSTLGAVSIG--NITSSSTQTRQTLTGLAARSDPDVAVNAQLKL--- 340
QY 1041 DDDNDADITTVGKDAKTQNOVNTLKLKGNGLDIOTKNDKGYTFYFGINOSGLKAGN--NTTL 1099
Db 341 -----AESGGGWNLTASGANSNVALBESVDLK--NSGNLITKTITDSNDVTRNLATL 394
QY 1100 NNNGLSINKTAGNEQIOVGADVGRFAKVNNGVYVAGIDGTTTRITFDEIGFAGTNGSLDKS 1159
Db 395 KYVDSLTTGNTA-----MTTDSGTIVGK-----RVLDSTGLVIAEG----- 429
QY 1160 KPHLSKDGINAGGKTIINOSGEIQAONSNDVATGKTIIDKTELENKISSAKTAQNSLH 1219
Db 430 -PSVISGINAGQKIMNNGTGA---DTDAVNFQGL-----QAVSDTA----- 469
QY 1220 EFSVADGEGNNTFVNPSSYDTSKTSQVITTFAGENGITTKVNGKGVVAGIDQTKGLTTP 1279
Db 470 ----- 469
QY 1280 KLVGNNGKGIIVDSONGONTITGLSNTLANVTNDKGSVTEGKIIKDEKTRAAST 1339

```

Db 470 -----SKGMNLLASGFTSSNVAVGASVDLKN-----TGNNLLI-----TKRIGI 508
 QY 1340 VDLVSLAGNLOGNEADVFSVTYTVAFADGNATTAKTATYDIDISKTYSVYDVVDDFTT 1399
 Db 509 NDVT---FNLATFALEA-DSLTTGTNTMTTDDVTGVGSVNT---LSTGTVLITD----- 553
 QY 1400 EVKDKKLGVKTTTLTSTGTGANKFALSNOATGDA---LYKASDIYVAHLNLTLS-GLDIOTAK 1455
 Db 554 -----GPSYSSSISAGNCKITINVAAGTADTAVNFSQLOAVSSTASKCMNLLAS 603
 QY 1456 GASQAN---SSAGYVDADGNKYVYDSTDNKKYQAKNDGTVDKTEKVAKDLVA--QAOPT 1509
 Db 604 GANSNNVVPGESVDLKNDSGNLITKTTS-----NDVFNLATALKVDSLTGTNTAMT 657
 QY 1510 PDGTLAOMNYSVINKEDVANDAK--KQGINEDNAFVKGLEKAASDNKTKNAATVVDLN 1567
 Db 658 TDGVTVGSNNVTLGSTGLVITDGPSTVSSGI-----SAGNCKITINVAAGTADTD 705
 QY 1568 AVAOTPLTFAGDGTGTAKKL-----GELTLTKGOTDTNKLTDNNIGV--VAGTDGFTVXL 1621
 Db 706 AVNFSQLOAVSSTASKCMNLLASGANSNVAVGESVDLKNITVTSKESGSDVLFNL 765
 QY 1622 AKDITNINSVNAAGTKIDDKGVSE---VDSGQAKANTPVLSANGLDLGKVIYVNGK 1676
 Db 766 SSSL-KIDKLTIVGDTVMTNGVTVSGVTLGSMGLVITDGPSTVSSGINAGSOKITNVAA 824
 QY 1677 GTKTDANAOQLNEVNNLGLGAGNADNADGNVNIADIKKDPNSGSSNNRYIYAKGV 1736
 Db 825 GTATDTDVNNLSQTL-----TA 840
 QY 1737 LGGKGNNDTEKLTAT--GGVQGVVDKDGNDGDLNVMVWKTOKOSKALLATYVNAAGTN 1794
 Db 841 MAGGASVNHYSYIDGDTQ-----GCNNGD----- 867
 QY 1795 YLNNPAPALDRINEGCIREFHVNDQNEPVOGRCIDSSASGKHSVAIGFOAKDGEA 1854
 Db 868 -----GATGTRSIAGVGTILASAG 887
 QY 1855 AVAIGRQTOA--GNQTAIGDNGAQTGDSIAIGT-----NVVAGKHS----- 1897
 Db 888 ATAVGSSAAASGKSTAGNAAVASADSVALDAGKARGAESYTGKSYGLONNTVGT 947
 QY 1898 -AIGDPSTVADNSYSVGNNOFTDAT---QTDVFG-----VGNNI-----TYESN 1940
 Db 948 VSVQDASKGFTRVSNVADAKKATDAVNLQDLRVADANRYDNKTESSEBQTEYKAN 1007
 QY 1941 SVALGSNA--ISAGTHA-----GTQAKKSDGTA-GTTTAGATGYVKGFGAGQTAVG-- 1989
 Db 1008 SL---NNSATPIIAGVADATAGVATASGADSIAMGKASASADNAV-----AIGNHS 1057
 QY 1990 -----AVSYGASGAERIRIQNAAGEVSATSTDAVNSGSLYKKTGCIANATNELLDRHIO 2043
 Db 1058 VADRANTVSGAGSEKQVTVNVAAG---TADTDAVNSQL---NQGILITAKQYTTDVVS 1111
 QY 2044 NENKANAGISSAMAMAMPQATYIGRSMVTGCIATHNGOGAVAVGLSKLSDNGQWFKIN 2103
 Db 1112 LRRDITDGVAAALATLANLPAYIIPGRMTSVGVSSYRGOSALIVGVSSVSESRWYFKFS 1171
 QY 2104 GSADTQGHGAVAGAFHF 2122
 Db 1172 GSANTRSQVIGIGAGVGYOW 1190

RESULT 5

AC0976
 Probable autotransporter sapB [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
 C:Species: Salmonella enterica subsp. enterica serovar Typhimurium
 A:Note: this species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
 C:Accession: AC0976
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churruarin, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S; Moutre, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churruarin, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moutre, S.; O'Gaora, P.
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AC0976
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1107 (CDS)
 A:Cross-references: GB:AL13382; PIDN:CAD03303.1; PID:G16504923; GSPDB:GN00176
 C:Genetics:
 A:Gene: sapB
 Query Match 6.3%; Score 670; DB 2; Length 1107;
 Best Local Similarity 23.7%; Pred. No. 1.4e-18;
 Matches 1304; Conservative 193; Mismatches 413; Indels 372; Gaps 59;
 QY 993 NSGDDKALLT-NAKDIADN---LNTLAGEIRNTKGA---DTALQTFQVKKYKENGDDN 1044
 Db 46 NAGNDTGSINQNTTDAITNTTTSINNLSNVTLTDLALMDASGTFASR---NGSASK 102
 QY 1045 DADTI--TWGDKAKTQNVN---TLKLGKNGLDIQTNKDGTVPFGINTQSGIKAGNN 1096
 Db 103 ITNLAAGTILADS-IDAVNGSQLPFTNEKVDQNTADITN---TNSINQNTTDLATNT 156
 QY 1097 TILNNGLSIKTACNEQIQVAGADGVKFAKVVNGVYVAGIDGTIRITDEIGFAGTNGSL 1156
 Db 157 TSINN-----LSNV-----TTLTDLALMDASGTF 183
 QY 1157 DSKPHLSKDGINAGSKRTITNIOGFIQNSNDAYTGKIVYDKLELNKISSTKATQON 1216
 Db 184 SASR-----NGSASKTITNLAAGTILADSIDAVNGSQLPFTNEKVDQNTADITN 234
 QY 1217 SLHFSVADQGNNTFTVSNPSSYDPTSKTSVYI-----TF-AGENGITTKV-NKGVYR 1267
 Db 235 SINQNTT-DIATNTTTSINNLSNVTLTDLALMDASGTFASRNSASRTNLAAGT 292
 QY 1268 VGIQDTKGLITTKLYVGNNGKGIYIDSON-GQNT--ITGSLNTLANVTNDKGSVRTTEQ 1324
 Db 293 LADSTDAV-----NGSQLYETNOKVDQNTSAIADINTSTINSSD----- 333
 QY 1325 GKLIKEDKTRASIVYDLVSAFNLQNGNEADVFSYTDYVAFADGNATTAKTATYDIDISK 1384
 Db 334 -----NLSMN-----ETTNSASAGSSSTTKITINVAAGE 363
 QY 1385 TSKVYDVNVDDTTEVAKDKLGKVTTLTSTGTGANKFALSNOATGDAVKAADYVAHL 1444
 Db 364 LSEESTDA-VNGSQLPFTNEKVDQNTADITN-----ITQNSTALENL 408
 QY 1445 NTLSDIQTAKASQANSAGVYDADGNKYVYDSTDNKYQAKNDGYDKTEKVAKDKLV 1504
 Db 409 NTSVSDINT-----SITGLTD---NALLMDE-DTGAFANHGSGTSKITNVA----- 451
 QY 1505 AOAOTPDGTLAOMNYSVINKEDVANDAKKQGINEDNAFVKGLEKAASDNKTKNAATVYV 1564
 Db 452 AALSDST-----DAVNSQLYETNOK-----VDQNTS-----AIA 483
 QY 1565 DINAVAQPLTFAGDGTGTAKKLGETTLIKGGQDTNKLTDNNIGVAVAGTGTGTVKLAKD 1624
 Db 484 DIN---TSITNLTGTDALSMWDEEGAFSAGTSGTNTN-----VAAQ-----ELASD 529
 QY 1625 LTNLNSVNAAGTKIDDKGV---SEYDSSGQAKANT--PVLSANGLDLGKVIYVNGKGT 1679
 Db 530 ST--DAIN--GSQLEYETNMLISQYNESISQLAGDTSETYITENG--TGKVIIRINDGLE 583
 QY 1680 DIDA-----ANVOQLNEVRNLLGNGNNDNADGNOVNIADIKKDPNSGSSNNRYIYAK 1733
 Db 584 GODATATNGATVAGYDVAAGAGCALQONSSSIBESIA-----LGSSTSRRAITTG 638
 QY 1734 -----GTVLGKGNNDTEKLTATGQVGVVDKDGNDGDLNVMVWKTOKOSKALL 1784
 Db 639 IRETATSIDGVI---GYNTTDRELLGALSIGT--DGESYQITNV-----ADGSE----- 684

QY 1785 ATYNAAGTNTLNNPAEALDRINEGIRFPHVNDGNEPVOGRNGIDSSAGKHSVAI 1844
 Db 685 -----ADAVTVBOLQNALGAVTTTPKYYHANSTEE-----DSLAVGTSJL 728
 QY 1845 GFOAKADEAVALIGRQTOA-----GNOSIAIDNQAOTGDSOIAIGTGNVAVGKHSAGI 1901
 Db 729 GAKTIVNADAGIGIGLNTLVNADALNGIAISNARAHANSIANGNSQITL-----RGKQTD 785
 QY 1902 PSTYKAD-----NSYVGNNN-----QFTD-----ATQDVEGVC-----NNITVT 1937
 Db 786 YTVANMDTPQNSVEFVSSEEDGQRLTNVAGSADPDVAVGQKVTDAQVSRNTQSIT 845
 QY 1938 ESN-----SVALGSNALSAGT-----HAGTQAKSDGTA-----CTTTAGATGVGFA 1993
 Db 846 NLNTOVSNLDRVTNIENGIDIVTGTSTKFKYKNTDQADNMAOGADSVAGSIAAAE 905
 QY 1984 GOTAVG-----AVSVGASGAERLIONVA-----2007
 Db 906 NSVALGTSVADDEANTVSVSSTQQRRTTNVAAAGVNNTDVAVNAQLKASEAGSVRYETNA 965
 QY 2008 -----GEVSA-----TSTDVAVNSQLYKATOGIANAIN-----ELDR 2040
 Db 966 DGSVNVSVLNGDGSQGTTRIGNVSAAVNDPDVAVYAOIKRSEAEANTYTDOKKGEMNSK 1025
 QY 2041 IHONENKMANAGISSAMAMASMPQAYIPGRSMVTGSIATHNOCAGAVVLSKLSNGQVVF 2100
 Db 1026 IKIENKMSGSIASAMAAALPQAPAGANMTSIAGTFENESAVALIGVSVSSEGGVNY 1085
 QY 2101 KINGSADTQGHVGAAGVAFHF 2122
 Db 1086 KLGTSNSGQDYSALIGAGFQW 1107

RESULT 6
 AD0123
 probable autotransporter protein yaph [imported] : Yersinia pestis (strain CO92)
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
 C:Accession: AD0123
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tiltball, R.W.; Holden, M.T.G.; Prentice, M.B.;
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AD0123
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-3705 <KUR>
 A:Cross-references: GB:AL590842; PTDN:CAC89847.1; PTD:915979073; GSPDB:GN00175
 C:Genetics:
 A:Gene: yaph

Query Match 6.2% Score 669; DB 2: Length 3705;
 Best Local Similarity 20.6%; Pred. No. 7.3e-18;
 Matches 593; Conservative 332; Mismatches 1003; Indels 962; Gaps 136;
 QY 7 VFNKATGTFMAVAYEAKSHSTGGSCATGQVSVRTLSEFARIALAVLIGATLNGSAY 66
 Db 365 VIYNGTGSIL-----KATNTGILATKNNNMSDI-YIRSA-----GDT 403
 QY 67 AGIGISEADGCGGANAKDKSIAIGDIAQALGSOSIA-----TGDKNIVHNSNNANIG 121
 Db 404 AATGISATGHTGTVKIKNDGTTSTTAGIAISASIKESIVDNTDGTITTAAGTGVNL 463
 QY 122 ARA-----SGNESIAIG-----DVLASGHSIAIGSDDLXL 153
 Db 464 ASAILNLEGGTINTSKANGITFAGTE-----GCHLLTDLTTLTGILSLSNVAGVNLTL 519
 QY 154 KREYVOISILLPIIRGKALNDIYOLADTNLQKRYRTHAOGHASTAVGAMSYAKGHFSN 213

Db 520 SNVTNLTN-----GVALNLTGLTLVDSLNGHNTINIEBAGIGIAATNTE-----LN 567
 QY 214 AFGTRATBEETSLAVGLIATA-----KAASSIANGNMAAIGRA-----ATAVGS 260
 Db 568 TFDPAELDINVNAGIGIGIQTGGVNLASNNLI-VVANTGLALQITDGIQNTTIGNE 626
 QY 261 TOYLNNGIA-----LGFOSQVLOKNDNVNANVAPAD-----295
 Db 627 IQLAENATRIINFLGSSSKLTNNNGTIGSVIFAGVADHI INNGNLTGDTLTGGAGNDL 686
 QY 296 -----DNQPI-----DNRYKATEKNGATDVEFSIGNSGNDSIRKLIINVA-----GSAD-- 340
 Db 687 VLDSSQSNQVNIINAGQNSVNTIONGAT-VSSITINGNDFTINGMSVSYLGSIDL 745
 QY 341 -----TAVVNAQIKEA-----VRLANROITKRGDS-----NNREKELGT 378
 Db 746 TGLNTLNFENASTDELAATSLQFTNINLVDSHITLVSDNIGSGMWNIDSSSELLFGST 805
 QY 379 -----LFTTGAQT-SALTDHNGV-----VONGDLKVOLETTLSLM-----417
 Db 806 FDGLIHAITLQAGTSAIVNNSANVSEQASMEFAGTWQVNGGALLTASNSQJASAKIGID 865
 QY 418 -----VTEN-----LANEKVTVKTRLTDDKIGFTNDNGIDESPRLDKOTGIHAGQ 468
 Db 866 GTLNLNDIALFNHVLFTNGTLNVAK-NLATTAFDFGSTVG-----GAFSGIV 911
 QY 469 KIRKLTAGVDDDAATYQOLK-KVNOFAESALQTF-TYKVKDKGN-----DAND 516
 Db 912 NLKRTTFALADNAALASATLKLSDSVTVGTDTDRTLHGLDLSGTLIFDCAVQSQOT 971
 QY 517 SKITTV-----GKNNKPDGTQVNTLK-----LKENGVDVTE-----549
 Db 972 SGVTVTVDLNLSGTVMITSSGMDNTPLATINVSITLEDORAGSTIELINATVTDIDA 1031
 QY 550 -----TNGT-VTFGLN-----QNNGLTVGNSYLNDCLSKYNINSKQIOY-----589
 Db 1032 LDLLVNGTATISGTQGVQSAIQOQGGSTVANA-IHNYGLASSNSNGSGLYVNYTSLALEL 1090
 QY 590 -----CAD-----GTFIDISNKP-GAG-----IENTRITL-----615
 Db 1091 LADADALLATESGLTANNVNLAEIPLFGVGLVYDAONGALTANSNRYEGTIVYAGE 1150
 QY 616 -----RDIGFANNTGSLDANKPRLPTGAGAGKEL 647
 Db 1151 LILGANGAFQOTSLILDASASANINQISQTVGAVTVGTVIGSGSVLTLGTLNGTL 1210
 QY 648 TNWQSAINPAT-----NGQIDFNNRLSTANTKSGSAATIKLILNL 669
 Db 1211 DLITGALNLTAGASVYAGGLTGAAGTLININGNLS-----VSANSGISGOT-----HIADV 1262
 QY 690 SQVPLTFAGDTPRVTKKEI-----LKKYGGKT-----ADDLTKNN--IGVVA 733
 Db 1263 ASVTLTDTGTLGTSAVEVLGTLNNGNAAAMTVNLSGDGTTINRAAVTLSGNNSFGAHQ 1322
 QY 734 DSTDNSLTVKLAKTLSDLDA-VNTKTLTA-----SDKV-----TVD-----768
 Db 1323 ICTGELVELVAGSNLGSASATVNLGTLFSLHLINGVESIANVLSGVAAGTVDIIGADT 1382
 QY 769 -----SGNNT-----AKLONDGLTF-SKONTGATP-----ATSKITIGVGLKFTDNGIALD 815
 Db 1383 ALTNANSGELQVYALAGNSKLTVASTNNLNGASSVALAGADTLISLGGFNGTGNSTGS 1442
 QY 816 GTTYITKD-KYFAKQDQ-----SLDKSPYLDKDKLKVGEVETTINGINAGKAITGLS 869
 Db 1443 GVLQVTDPAEVLTLSSNGVSAVYITIDATDLNLDIALNRHVLTGNGL-----LN 1493
 QY 870 NTLDTATNA-----TTGHTVQLGIVDSTDKTRAASISDVNL-AGFNILKNGDAKDEPVSY 923
 Db 1494 VAKNDASAPFEGSTVGAGFS-GIVNLTNTTFALSAUNAAALARATKLSDSVTVYGAT 1552
 QY 924 D-TYDFINGNATTAKVYIDKASK-----VADVAVDGTHTLITLQADG--KKNQGVK 973
 Db 1553 DRLTGLDNLNGT--LIFDGPPOSQANGVTVYIDLALNSGTISITGAGMWEHHPVTPP 1610

QY 974 TTTTCTKADG-----KAINFSVSGDDKALINAKDIADMLNLTAGEIRNTKGTADTA 1027
 Db 1611 NVSLLEBOD-RGDILLELINANVTGNANLLDLDCTALTSCGYESAIQGGSTVANA 1669
 QY 1028 LQTFQKRYKENDDD-----NDADITTVCKDA--KTNOYNTIKLKGKNG 1071
 Db 1670 IHNYGLTSSNGSGSLVYNTLSALELLANGANALLLATESGLANRYLMELEFGVGL 1729
 QY 1072 DIOTNR-----DGTAT-----FG-----INQSGKAGNN----- 1096
 Db 1730 VYDQNGALTLANGNNRRECTITTYAGELLGANGAFQOTSLINTASGSANINGRQTY 1789
 QY 1097 -----TTLNNGSLIKN-----TAGNEQIQVADGVKFAKVNNG 1130
 Db 1790 GAVTNGAVTLGNGVLTGSLTNGGILDLTGALNLAAAGSSVYAGLTGAGTLINNG 1849
 QY 1131 -----VVGAGIDGTTIR-DEIGFAGTNGSLDKRP-----LSKDGITAGKKTITNIO 1180
 Db 1850 DLAVSAFNGSLGSGQTHIADVASVTLGT-GTLGTSAVEVLGTLNLTNGANA--MTNVL 1905
 QY 1181 GEIAQNSNDVATGKTYLDEKTELENKISSTAKTAQNSLHEFSVADQGN--FTVSNPYS 1238
 Db 1906 GGGVITMAAVT-----LSGNNS--FGARIGTIDGELTVGQASN 1943
 QY 1239 SYDTSKSDVITPAGE--NGIT--TKVNGVVRVGIQDTKGLAT----- 1278
 Db 1944 LGASSATVNLGTLTSHLILNGVSESIANVLGAGSTVDIIGADPALTANNSGFLGQYA 2003
 QY 1279 ----PRUYGNNNGK-----IYDSONGNTITGLSNTLANTNDKGSRTTEQKIID 1330
 Db 2004 LAGNSLTVASTNNLGAASSVALAGADTLISLGFNGTGNVSGVQLQVYDADAEVTLT 2063
 QY 1331 EDKTRAASI-VDLVLSAGFNLQ-----GNG-----EAVDFVSTY-----DT 1364
 Db 2064 SSNGVGNVYKVDIADATLNLNDIALEFDHVLGTNGTLNVAKNLATAFDPGSGVGAFFSI 2123
 QY 1365 VNF-----ADGNATYAKVY--DGT-----SKTSKVVYDVNDTT----- 1398
 Db 2124 VNLNTTFALSDNMAALRAATLKLSDSDVTGVTDRILHGLDLNGTLIFDGSPPQSG 2183
 QY 1399 ----IEVKKKKLGVKTTTLTSTGTGANK-----FALSNOATGD--ALVVASDIAVHL 1444
 Db 2184 ANGCVITVTDIALNSGISTGANGMENEHPVPVNSLLEQORGLDQLIDADVTGTA 2243
 QY 1445 NTL-----SGDIQTAKGASOANSSAGVYADGNKVLYDSTDNKYQAOKNDGT---VDKTK 1496
 Db 2244 NDELMINCTTISAGGVSTVQGGGYTANAF-----HNYGTSNGSGGLVYNT- 2294
 QY 1497 EVAKDKLVAAQOTPDGTLAQMNVKSYINKEQVNDANKKOGINEDNAFYKGLKASDNKT 1556
 Db 2295 ----LSALELLADGA-----NALLATESGLTANRE 2321
 QY 1557 KNAAYT-VGDLNAAVOT-PLTEA-----GDTGTAKKL-----GET--LTIKGGQ 1597
 Db 2322 LNAELSGVGLVYADQNGALTLANGNNRREGTITVYAGELLGANGAFQOTSLINTASGA 2381
 QY 1598 TDTNKLTDNNGVAGTDFYVLAKLDTNLNS--VNAAGTKIDKGVSVSDSQAOK-- 1653
 Db 2382 SANINGRQTVGAVNTGTGVLNGGELTSTDTLLNTGMINTD-GIINLENGGASSISG 2440
 QY 1654 ----ANTPVLSANGIDLGKAYISVNGK--TKDTPAANVQOINEVRNLGLGNAGNADNAQ 1710
 Db 2441 GLTNGILNIKGGDFTISIDNNGLAGOTNISGASV-TLNGCTIIGTGNGLS----- 2492
 QY 1711 VNIAIDKPPNSGSSSRVIVKAGVYLGKGNNDTEKLT-----GGVYGVYDKDGN 1762
 Db 2493 -SVIVDLGLNLVADNS-----LANVISGDGTINTTATVYTLGSSNSFGAHQIG----- 2540
 QY 1763 ANGDSLNVVVKQDKGSKKALLATYNAAGQTYL--TNNEPAEADRI----- 1807
 Db 2541 TNGELT-----VGQASNLGASSATVNLGTLTSHLILNGVSESIANVLGAGSTVDIIG 2595

QY 1808 -----NEGIRFEFHVNDNGQEPVQGRNGIDSSAS-----GKHSVAIGF----- 1846
 Db 2596 ADTALTANNSGFLQYALAGNSKLTVASTNNLGAASSVALAGTDTLSLGFNCTFGNSV 2655
 QY 1847 -----QAKADGEA-----VAIGQTOAGNSAIGNAQ 1876
 Db 2656 TGGSVLQVTDAAEYTLTSSNGSVNAVTIADATLNDLIDTALFHHALTGGLLVAN-- 2713
 QY 1877 ATGDOSIAIGTGNVAGKHSAGIDPSTVKADNS-YGVGNNOPTDATTOT-----DVFG 1929
 Db 2714 --DASTAFDRGATVGAFTG-----TVNLNSTFPDLSGNTTYVALATLKLSSGNTLS 2764
 QY 1930 VGNNT-----VYESNSVALGSNSAISAG--THACTQAKKSDGTAG----- 1968
 Db 2765 VGNQVONIGTILAMNGCTLFPNIVDNGILTSDDGTIANSTINTGGEVRYNLPNSLAPS 2824
 QY 1969 -----TTTGATGATG-----VKFAGQTAGVANSVAGS 1997
 Db 2835 LDGLSMELDECEITIVLTATGAATGTGHELTLDENGDPISAVTVQGVHNGASTSAATG 2884
 QY 1998 AERR-----IQNVAGEVSATSTDAV-----NGSOLYK----- 2025
 Db 2885 SPNYGMTTGEDYDGLVYVYVGLTLELSTGSEALVTLAAILANNQTSNDLSAQITSGDL 2944
 QY 2026 -----ATGCIANATNEL-----DRIHQENRANAGISSAMAMSPQATYIPGRS 2070
 Db 2945 AFASANDSTASLNSNSTSYTGTTWASSGNRLDADSA-LGQTSILAMSTATHVDINGTQ 3003
 QY 2071 MYTGIAATHNG-----OGAVAVGLSKLSDNQWFKRINGSADTOGHV 2112
 Db 3004 QVGEELATEGSSSTLDLNDKILTVTGGOIDGALVGGGELVLSG--LNVSTDNAGFT 3059
 QY 2113 GA--AVAGCFH 2121
 Db 3060 GSTDIANGAVAH 3071
 RESULT 7
 B11704
 cell surface antigen (sca3) RP451 - Rickettsia prowazekii
 C:Species: Rickettsia prowazekii
 C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
 C:Accession: B11704
 R:Anderson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark
 Nature 396, 133-140, 1998
 A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
 A:Reference number: A71630; MUID:99039499
 A:Accession: B11704
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2340 <AND>
 A:Cross-references: GB:AJ235271; GB:AJ235269; MID:g3868717; PIDN:CA14908.1; PID:g386
 A:Experimental source: strain Madrid E
 C:Genetics:
 A:Gene: sca3; RP451
 Query Match 6.1%; Score 651.5; DB 2; Length 2340;
 Best Local Similarity 21.2%; Pred. No. 1.9e-17;
 Matches 519; Conservative 302; Mismatches 879; Indels 743; Gaps 121;
 QY 10 NKATGPMNAVE-----YANSHSTG-----GGSCA-----TGQVSVFTLSFARIALAVL 55
 Db 4 SKIIRKFLATASLGTLFTNSMATGTIIPNNGSVSLNTDAGLVGV-----FNNGDITQI- 58
 QY 56 VIGATNGSAAVAGIGIEADGCRGAGANARDKSAIDIDNALGOSAIAT----- 105
 Db 59 -----VWGRIKISADKANAIIGTINTLKE-----LPDFGGEVSEYQVNSVIGPLNAGEDLN 109
 QY 106 --GDNKIVHS-----NNNANIGAK-----ASGNEST 130
 Db 110 TNFPLPLFISNNVSIITGVGTRTFNSIDFAGKNATIQINKDLNITTKIDNTVAGNNGSI 169

| | | | | |
|----|------|---|---|---------------------------|
| 0Y | 131 | AIGDVLASH----- | ASIAIGSDULYKREYQOISELLPIITROKALINDYOLADT | 183 |
| Db | 170 | TEBESSIGLISHNIEYTNLSLGINNGEAKIYAPEANNT | -----I | 209 |
| 0Y | 184 | NLOQYRTHAOGHASTAVGAMSYAKHFENAGCTRATAGEYSLAVGLTATAKASSIAV | 243 | |
| Db | 210 | NAKININTLTHNNSLITLDCDNIITTLAKONINNT | ---TEIDOGILINLAYD----- | GSSSIIT 262 |
| 0Y | 244 | GSMNOAIGFAATAVGSGTOYNLNRGIALGFGSOVLQKDNVMAANVARYADDNOPIDNR | 303 | |
| Db | 263 | GD-----IG----- | NIGSLDITNVLGGSAT-FNSTILKATNINLKHFTSLTLDN----- | 307 |
| 0Y | 304 | YKATFKKGAADVSGISNGSNDISIRKIIIN | ---VGAGSADPTDAVNAVQOLKEAVRLANOIT | 361 |
| Db | 308 | ----- | IIVIGNIGNNN--KDLINFRHGHTINDNEMIIIPAPK----- | THGTLN 349 |
| 0Y | 362 | FKGDGSMN----- | RYEKGIGKTLITTTGAO----- | TSA 389 |
| Db | 350 | FKGMAITLGNINNLINLIKPSGGHGKTLNLQGTTRYKDNLYFADSVLDSGTISVNLDTDC | 409 | |
| 0Y | 390 | LTDHNIGVONGDLKAVQLAETL----- | TSIKVTEENTLANEKVTGKTRTLTTDKIG | 442 |
| Db | 410 | VTFNNSNV--NGGTLIIAKNTISAKLINAATKAKIQUANALTMNH----- | ----- | PSAG 456 |
| 0Y | 443 | FTNDMNGIDESKPRLDKDCIHAHGCKITKLAGVDDDAATGOLKRVNQTASALQTF | 502 | |
| Db | 457 | DISIRIADTIYITIDAKN----- | GNVNLNNNAKIIIFEGADS--MLALIN--TGVTADRTF | 509 |
| 0Y | 503 | TV-KKYVRKNGNDAND----- | SKITTVKRNKPKPGT-OVNTLKLK----- | 539 |
| Db | 510 | TIYNLNLQSGNDEIGYIKLEAIKKVITFINQSGPTTIGODNTHRKLEILVAGADIIID | 569 | |
| 0Y | 540 | ----- | GENGVDTTETNGVTTFGLNQNNGLTVGN | 568 |
| Db | 570 | TIIFKLINSTGTFPNRPLDLAGAGNIAFGHGLTVNGVTSITISENNQGLIFNS | 629 | |
| 0Y | 569 | STLNDGLSKYKNTNSKQIOVGADGIT----- | FTDISNSKPGA----- | GI----- 608 |
| Db | 630 | G--NITGVTIGNEGLKLVINIGADPYTCANAFASALNPSVILIADGVTLGEVYTH | 687 | |
| 0Y | 609 | ENTRITRDRDIGFANNNGTSLDANKPRLPTPGINAGGKELTNVQSIINPATNGGOLDFMNR | 668 | |
| Db | 668 | NNTGCVLSLGTG--SNITGQIGTNSAALKEINIGAGA-- | SNIDSNIYAG----- | 732 |
| 0Y | 669 | LSTANTEKSGSAATIKDLYMISOVPLTFPAGDPNPTKKLIGELILVKKGKTTA--DOLTKN | 727 | |
| Db | 733 | STYLTJDOTSELPLMNDVYVNSNIITTAGNSGSKILFTNGGTCGIDANANGALOEVYFN | 791 | |
| 0Y | 728 | ---NIGVADSTDNLSLTVKLAK----- | TLSDDAVNTKTLTASDKV--VDSGNNPAKL | 776 |
| Db | 792 | GTTNIGGFANSOMFTYVAHSANVITGLTTFGALKYKIDCTIIAHGVLGVIDIDFNNAKGKF | 851 | |
| 0Y | 777 | QNGD----- | LTF----- | SKONTGATPNTSKTIG----- 800 |
| Db | 852 | TLIGDAMIDGVLONGGVACTLDF | IGDGNVTONIGADNANSSTINIOGDNTKNVTIAND | 911 |
| 0Y | 801 | --VDGLFTDNDNGIADGTYITIKRVGVGAKODSGIDSKPYLKDKLKVEVEITTINGI | 858 | |
| Db | 912 | IFVDNIHTNGIQLQLOGN--LTHNHIDGANGGLEEFN----- | GNNTYINLAI | 958 |
| 0Y | 859 | NAGGKAITGSLNTLTD--ATNATG----- | HYTQIGI----- | VDSDTKTPAASIGVL 904 |
| Db | 959 | IVNGO--NGILNATFNKASDDTIGVTKIINIQTISTPONFTIOVNNKMLTIVSSVSSI | 1016 | |
| 0Y | 905 | NAGFNLKANGAKD----- | FVSTDYDVEJING--NATKATYTGKRSKVAAYVNDGTIIH | 959 |
| Db | 1017 | NF----- | GSANSQILSLAPVDQTIKFINNLNETGGG-- | TLTDSN--GNMLT 10599 |
| 0Y | 960 | LTGADGNKNOJGVTTTLTFTDAKGDKAI----- | NFVNSG----- | DDKALINADI 1006 |
| Db | 1060 | ISGNNGIK--IGSGNSELSSIMIKGVYTIINDLIONHQLMINNGALFDDQSLSAIKI | 1117 | |
| 0Y | 1007 | ADNLMTLAGIIRNTKGTADPALQTFQVKKVKEGDDNDN | ----- | DTTVGKDAKTNOV 1060 |

| | | | | | | |
|----|------|---|---|--|---------|------|
| Qy | 1118 | NINIGIYAG- | ----- | ATYLLDIAINDPDLINTSGWFEKHQDIDILELKSNTND | 1165 | |
| Qy | 1061 | NTELKAKCKNGJDIQTAKDGTWTFGINTOSGLKAGNNTLLNNGLSIKRTAG- | NEQIOVG | 1118 | | |
| Db | 1166 | HTITL--TSALDPDGNNOFGIILKITDT | ----- | NKLTIDNG-NVATYLGANMLK- | 1213 | |
| Qy | 1119 | ADGVKFAKVVNGV---- | YGAGIDGTRITRD-EIGFAGINGSLOKSKPHLSKOGINA---- | 1170 | | |
| Db | 1214 | -QTFEASIDINGIAKLVGINTENVTLLNKDIELEVANANVLENTTYTATGNIHVD | 1271 | | | |
| Qy | 1171 | -GGKRTNIQ----- | SGEIAQSNDAVGT----- | GKIDYDKTELENNKISSAKTAON | 1216 | |
| Db | 1272 | FQAGAYIINLDIEIDGVSYTGNC- | WGTILNFGSGKYGLINNIYMLQAGADYSL | 1329 | | |
| Qy | 1217 | SLHEFSYADIG-- | NNFT-VSNPYSSYDTSKT----- | SD | 1247 | |
| Db | 1330 | ASGVNSTEIOGNGNNLTPAANSHTTIDINKTGODLNVFINGCVSGSIGANAAYD | 1389 | | | |
| Qy | 1248 | VYTFAG----- | ENGITTKVKNKGVYRGD-QRKGLTTPKL----- | 1281 | | |
| Db | 1390 | IINAGVNSNTLKSQNIYISGATMOYNNVTATDJSKNANNGTILKNHPINITS | 1449 | | | |
| Qy | 1282 | TVGNNNGKGIYDSONQONTIT----- | GSNTLANVYTDKGSVR | 1320 | | |
| Db | 1450 | TLGNNNAIG-TIEVANNDVTITSTLOAQNIHFSNATQAATTLGAAISOYTITTAGNNIH | 1508 | | | |
| Qy | 1321 | TTEGKRIKEDKTRAAIYDVLS- | AGENLONG----- | EAVDFYSTDYTVFADGNA- | 1372 | |
| Db | 1509 | TLE----- | VTPEIDGNCIIODDANNRLKSEILNGVGTINSHVYSSITTLANNQCNKL | 1564 | | |
| Qy | 1373 | -TTAKVTYDDTSKTSKVVYDVNDFTT- | EYKDKKLVGKTTTLSTGTANKPALSQA | 1429 | | |
| Db | 1565 | NIEGITYDDGSKTKSLANOISEPTITGVDYSKYLIINDAKTINFPDGNMNMKMLD | 1624 | | | |
| Qy | 1430 | TGDMVY----- | KASDIYAHNLNLSD-QZPAKGAQAN-SSAGVYVDD-GNKVYIYSTD | 1480 | | |
| Db | 1625 | IPDALIDLVLPSLSLFNFETDIKADNLNFAADGTAFNAKRAYVIDAHIDNGILKPN | 1664 | | | |
| Qy | 1481 | NKY--YOAKNDGTVDKTEFEYAKDKLVAQAOTPDGTLOAQNVS----- | VINKEQ | 1527 | | |
| Db | 1665 | NAMLTQELKANIT----- | ELASQEM----- | LLOKNIKAAULLIDMNLVLLDNVE | 1721 | |
| Qy | 1528 | VNDANKKQGINEDNAPYKLEKASDNKTKNAAVTYGDLNAAQPLTPAGDGTIAKL | 1587 | | | |
| Db | 1732 | VNTMLNVDIYDLDA-- | NYELKTYGNTVNHGILIT----- | ITY-- | FDALQKG | 1775 |
| Qy | 1568 | GETLTINGGQTDNKLTDNNIGVYAGTDGTYVLAKDILNLS----- | WAGETK | 1637 | | |
| Db | 1776 | GHIIVSGSNVMSDL-DNLI----- | IKTASHDITNITSYDKHQIVKLETGALTP | 1826 | | |
| Qy | 1638 | IDDKGVFVDSGQAKANTPYLSANG- | DLGKVISNVG----- | KGTKTDPAANYQ | 1688 | |
| Db | 1827 | VPOKVIITIDASEQNKVKKVVAANGVLLTIDGGR- | DDTGGRDITGRGRTD----- | 1878 | | |
| Qy | 1669 | LENEVRNLGLGN----- | AGDNADGNOVNITDIK-- | DPN----- | SGSS | 1726 |
| Db | 1879 | -NGCRDMCDVAGNISNNSNAGSSSPDN-YGITVDVPIFDESSPLIDYTKNNVYASGIAN | 1936 | | | |
| Qy | 1727 | --KRYIAKQTYGVGCGNNDTEKLATGCVQVVDKG-- | NANG----- | DLSWVWK- | 1773 | |
| Db | 1937 | QILNHVDFEGTDTAGLNDLIDFMSRPNRTETLDRLSNRIVANGINBEGVGLNIEVEN | 1996 | | | |
| Qy | 1774 | ----- | TOKD-GSKKALLA-- | TYNAGQNTYLTNNPAALIRINEOGITREFHV | 1817 | |
| Db | 1997 | FLTDIAINMDFETAKELGNLEELSDANVYNGIKNTJLNNK----- | INLKL | 2045 | | |
| Qy | 1818 | KDQNOEPVQGRNIDSSASGKSHVALIGFOAKADGEAAVAGROTQAGNOSTAIGDNRQA | 1877 | | | |
| Db | 2046 | NTNNOAIIIAAGDE-- | DNIVYIGWMSYFKIKONSKN-- | ASGYQNTGGGIIIGEFYNI- | 2100 | |
| Qy | 1878 | TGDSIAIGCNVYAG-- | KHSGAIGDPYVKAADNSYSVG----- | NNNQFTDA----- | 1922 | |

Db 2101 --DNSIVGAATYTMADSKVHKNDKGRTRKAKSNISYIGLYNMLTNFVEAIGVGR 2158
QY 1923 -----TQDYFGVGNNTITVESNSVALGSSAISACT 1954
Db 2159 NKIKYERKRTITTDQIALGKFINFTFYELLGIVNLTISHRT 2201

RESULT 8

hypothetical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain R)

C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F90696

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasaara, N.; Yasunaga, T.; Kihara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F90696

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-5291 <HAY>
A:Cross-References: GB:BA000007; PIDN:BA03965.1; PID:913360000; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952

C:Genetics:
A:Gene: ECs0542

Query Match 6.0%; Score 645.5; DB 2; Length 5291;
Best Local Similarity 21.9%; Pred. No. 9.4e-17;
Matches 582; Conservative 293; Mismatches 932; Indels 855; Gaps 131;

QY 35 TCGOVSVRFLSFARIALAFLVIGATLN---GSAVAGIGISEADGG-KGANARGDKSI 89
Db 1136 SGPFTLTINTVSGDDIINAEIYVAOTISQVGTGTAAGTIVYTGGMQYATVGSDDL 1195
QY 90 AI---GDAIALGQSOSIALGDKRIYHNSNNANIGAKSGNESIALGGDVLAGHASTAI 146
Db 1196 SVSVPANVLOALGCELFTISAS--LTNSANN---TGTAHVIDIDANLPGRLVPTV 1246
QY 147 GSDDLATKKEVVOISELLPIIRGOKALNDIYOLADPTNLOKRRRHAGHASTAY---- 201
Db 1247 AGDDVYNSIETHTQAL---VITGSSS---GLAAGALTVYINSVT 1284
QY 202 -GAMSTAKGFHFNACFTRATF---GTSLAVGLATATAKASSIANGVAOIGFATATV 257
Db 1285 YGATVLAOSMSGVADVADVTWMPACTVNIANGTNTACTTTSI---SHPTVVDLAAYAI 1341
QY 258 -----GSGTOYVNLRGIALGFGSQ---VLQKDN---DYNA 286
Db 1342 TINTLTDDVINAEEKSDQLSGLTSGVEAGQITVIFGSKSYTTVAADNTMGLTIPA 1401
QY 287 ANVRATVPDDNPIDNRKATFKNGATV-FSIGNSNGDSIRKRIINVGAS----- 338
Db 1402 VDV-----ATLPDGAANYOASVSNAGSTQATHAYSVDAVAPSVTINT 1445
QY 339 -ADTVAVNAQLKEAVRLANRQ-----ITFGDDSNRVEKGLTGLTITGG----- 384
Db 1446 IATFDILNAEAGSALTITGTAEGQIVYTLNGVANSNGVADGMSVSVPFGDILAS 1505
QY 385 -----AQISALFDHNIYVONGDGLKVOLETLLSKMVTENTL----- 423
Db 1506 LPAASYTVNASVSDKARNSASATHN-----LTVDLAAPVYVINTVACDDIINATEH 1556
QY 424 -----TANE-KATVGTGRLTTDKIGFTNOMNGIDESKPYLDKDTGIGHAGQ 468
Db 1557 GQAQITISGATCATGTGNTVSVITGTYTT-----VLDAAG-----NMSIGVPAS 1601
QY 469 KITKTLTAG-----VDDDAATYGOALKKVNQTAES-----ALQTFYKKVKDKNANDSK 518
Db 1602 VTSALAGGVITTAATYTDGANGSGTASHIVYALGAPVLAITIAVDIINAEAGDALA 1661
QY 519 ITTVGKNNKPDGTQVNTLKLKGENGVDTTETNG---TVTFGLNONGL-----TVGNSL 571

Db 1662 I--GTSNQFAGTQI--TVTLNGON--YTTTADAGSNMVSVPASVSLGEATVYTAAT 1717
QY 572 NNDGLSYAKNINSKQIOVADGDTF-----TD-----ISKRGAGIENTTRI 614
Db 1718 DADGNS--GSASHNVQVNTALPGVTINVAATDDIINAEAGEEQIISQVYGAAGDITV 1776
QY 615 TRDGIQFANNNGSLDANKPRLPTGINA-----GGEKELTNQSAINPATNG----- 661
Db 1777 T---LGCATYATVQANISMSVDVPASALDELNGELLISASVTNSGNTGNGTRETID 1833
QY 662 -----QIDFNKRSSTANTKESGSAAT---KDLVNLQVPLTFAGDT----- 700
Db 1834 ANLPLGRVDYAGDVNIIIEHGAIVTSSSSGLAAGSVNTLLINQTYAVALADGTW 1893
QY 701 ---GPNYTK---LGEIILKKGKTYAD-----DLTKNNT---GVADSDTNSL 740
Db 1894 SVGPADVSAWPAQSVYTIASGSGTSAGNPSVYTHPTVLDLSAVALSINAITADDTVINA 1953
QY 741 TVKLAKTLSDLDVANTKTLTASDKVTVDGNN-----AKLQNGD 780
Db 1954 EKGALTLISG---STSGVEAGQTVTVTFGKITYSATVAANGSKTSVPAADMAALRDGD 2009
QY 781 LT---FSKONGATPATNSKTIQVD-----GLKFTDNN----- 810
Db 2010 ASAGASVSNVNGSATTTTHAVSDASAPVTYINTIAGDDILINAEAGALTTGSSSTAE 2069
QY 811 ---GIALDCTTYTTRKQVFRKADGSLDKSKPLDKL-----KVG----- 849
Db 2070 GQTVTVTLNGTNY---TGTVQDGSWSVSPADSLTASNYTVNAVSPADKGPAS 2124
QY 850 -----EVEITTINGINGKAITGLSNTL---DPTNATGHTVTOIGIVDT--- 892
Db 2125 VNHMLTDTSVPTVINTVAGDDVYINATEHQAQIISGATGATGATVYTGITNFTT 2184
QY 893 -DKTPAASIG-----DVLNAGFLKNNGDAKDFVS-----TYD 924
Db 2185 VLDSAGMSVGPASVVSALANGTVTINASTVDAG---GNSGKHQVYNTGLPTTFEN 2241
QY 925 IV---DFINGNATYAKTYDQKASKAVADVNDVSTTHLNG-----ADGKNNQGVK 973
Db 2242 AISDNIIMADEKQPLTISGSGTGLTAQV---VTLNGHNSATTTDASGN-----W 2292
QY 974 TTTTLTKDAGDKAKINFSVNSGGDKALINAKDIADNT-----NTLAGE--IRNT 1020
Db 2293 TTVPVSDIALAGQANTVVSATSAAGNTASSANLTVDSGLPDTYINTVAGDDIINAA 2352
QY 1021 KGTADTALQTFQVKKVKEGDDNDADDTTVGKADAKTNOYNT-----LKLKG 1068
Db 2353 EAGADQITISGVTPRAA-----GDTVTVLGGNTYATVQSNLSMSVSVPADLQALGN 2406
QY 1069 NGDLIO---TFKDGTVFNGINTOSGLKAGNNTLNNN--GLSINKRANGQIO----- 1116
Db 2407 GDLTITASVYINANG-----NTGSGTR--DITIDALPGLRVYVAGDDIVSIEHQA 2457
QY 1117 -----VGADEVKFAKVNNGVAGAGI--DGTTRI-----TRDEIGFA 1150
Db 2458 LVITGSSGLNAGAVLVITINSVAISATVQADGWSVGIIPAANVSAPAGLVEVDGQS 2517
QY 1151 GTNGLDKSKPH-----LSKDCINAGKKTINIOSGELIAQ--NSMDAVT--- 1192
Db 2518 SANNPVSHPFTVLDLRAVAISINTVASDDVINAEEKGMLTISGSGIESQQTIVTF 2577
QY 1193 GKKIY-----DLKT-----ELENNISSTATAONSLHEFSVADGQ 1229
Db 2578 GKKIYTVASVAANGSWSNVPAADLATTPEGAANYOASVSASGNSASATHAISV--DASAP 2636
QY 1230 NFTVSNPYSSYD---TSKTSDVITFAG---ENGITKVN--KGVVRGIDOTKG---LTT 1278
Db 2637 TLTIT--NTIASDDILINAEAGSPLTISGTAHVGQVTVTLNATYTGVOAGMSVS 2695
QY 1279 PKLTVGNNGKGIYD---SONGONTTIGLSNTLANVTNDKGSVRTTEGKTIKDEDKTR 1335


```

Db 2696 PTSLGALNASTVTSATVNDKAGNPGSASHNLAVDTTAPVLTINTVAGDDIINDAEHAQ 2755
QY 1336 AASTVDLVSAGFNQNGNEADVPVS-----TYOTVNNADONATTAKTYYDDTSEKTSKV 1389
Db 2756 AL-VISGTSAG-----GEAGDVVSVVLNGKTYTTTLDASGNWS-----2792
QY 1390 YDVAVDDTTIEVKDKKLKTTLTSTGTA-----NKFALSNQATGALVKASDIY 1441
Db 2793 -----VVPADVYALGSGAQTITASVSDAGNSDDASRTVYVLSLAPV 2836
QY 1442 AHLTLDSD---IQAKGASGANSAGYVDAGNKVYIDSTDNKYQAOKNDGVDTKEV 1498
Db 2837 ISINTIAGDDVYINATEKSGDLALS--GTSDDPAGATVITVINGQVYSATTDASGWNVSIV 2894
QY 1499 AKDKVLAQAQTPDGLAQMANKSVYLNKEQVNDANKKQGINEDNAFVKGLEAAADNKTKN 1558
Db 2895 P-----ASAVSALGEATVSVTASVYN-----AQNSTASHNVOVN 2930
QY 1559 AAVTVGDLNAAVQPLTFAGDGTGTTAKKLGFTLIKGGQDTNKLTDNNIGVAGTDGFT 1618
Db 2931 TALPBITTPNPAATDDIINASEGSAQTISGQVTCMAAGSTYVEL-----GKRTYT 2981
QY 1619 VKLAKDLTNLSVYNGAGTKIDKGVSEVDSGQAKANTPVLSANGLDGKRVISNWKGT 1678
Db 2982 ATVQADLSMNSVVPADWQ-----ALNGELTVNASVYNA-----VGNISGSGT 3024
QY 1679 KD--TDANVQOOLNEVRLLGL--GNAGNDNAGNOYVNIAD-1KKDPNSGSSSKRYIYK 1732
Db 3025 RDTITDA-----SLPGLRVDYVAGD-----VYNIIEHQAQVITSSSG--FA 3066
QY 1733 AGTVLGKGNNDT---EKLATGVQVGVKDKGNANGDLSNWKVQKQKSGKALLATYNA 1789
Db 3067 AGTALTIVINNQTATVYTLVANGMSVGPAT-----DVSN-W-----PAGTIANITYSGANS 3116
QY 1790 AGOTVYLN-----NPAEADRIN--EAGIRF-----FHYNDGNEPVVQ 1828
Db 3117 AGTQSTHPLTVDLTAVALISNISTSDAIINAEKGAALTLSSGSEVAGQVTVTFEG 3176
QY 1829 RAGIDS---SASRKHV---AIGFOKADGEAAVAIGROTQAGNOSIAIGDAAQATGDOST 1883
Db 3177 GKTYYTTVAANSMSTVPADLAALRDGDASAOV-RVTNNV-----GNSATATHEYSV 3229
QY 1884 -----AIGTNVVAAGKHSGA---IGDPTVKADNSVY---GANNQFETATODV 1927
Db 3230 DSAAPVYIINTIASDNIINASPAAGVYSGTSTAQTGTLVTLNGYVOTT--VQTD- 3286
QY 1928 FGVGN-NITVYESNSVALGSNSAISAGT---HAGTOAKKSDG-TAGT-----TTAG- 1974
Db 3287 ---GMSLTLPRASDLTALANNGYTLTATVSDLAGNLGSASGVYDTTAPVISEFTVAGD 3343
QY 1975 -----ATGVYKG-----TAGQ-----TAVGAVSYGASGAERIONV 2005
Db 3344 DVINNVHEIOAIIISGTATGAAGDRLVYTIAGQOYVITSDASGMSVGPASV--ISGL 3401
QY 2006 AAGEV--SATSTDAVNGS-----OLYKATGIGIANATNEIDHRIHONENKANAGISSAM 2056
Db 3402 ADGTYVTISATTTDSAGNSSTOCHNOVNTAVALSVSTISGDNLA---MAEAG--SAL 3455
QY 2057 AASAPROATVPGRSWVTGCIAT--HNGOGAAVAVGLSKLSDNGQWFKI-----2102
Db 3456 TISG-----TGTNFRATGVYTVLLNGKYSAT-----IOSNSMSVNPADVAALSDGT 3505
QY 2103 -----NGSADTQGH 2111
Db 3506 SYTVSASADSGSAGNGNSTOTH 3527

```

```

R:Perna, N.T.; Plunkett, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409: 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MIMD:21074935; PMID:11206551
A:Accession: B85547
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-5188 <STO>
A:Cross-references: GB:AE005174; NID:g12513366; PIDN:AA654838.1; GSPDB:GN00145; UMGF:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 20615

Query Match 5.9%; Score 634; DB 2; Length 5188;
Best Local Similarity: 22.0%; Pred. No. 2, 6e-16;
Matches 576; Conservative 290; Mismatches 928; Indels 830; Gaps 130;

QY 35 TGGVGSVRTISFARIALAVLIGATLN-----GSAYAGIGISEADGG-KGANAARDKSI 89
Db 1136 SGPLTINTVSGDDIINAEIYVAQTIISGQVGTAVAGNTVYITGNGYNATVQSDLSM 1195
QY 90 AI---GDIAQALGQSIAIGDNKIHNNSNNNANIGAKASGNEIAIGDYLASGHASIAI 146
Db 1196 SVSVPANVLQALGELTISAS--LTNSANN-----TGTATHDIYIDANLPGLRDYV 1246
QY 147 GSDDLVKKETVQOISLPLIRGOKALNDIYOLADTNLQKRYTHAOGASTAV-----201
Db 1247 AGDDVINSIEHTQAL-----VITGSSS-----GLAGGALTVVINSVT 1284
QY 202 -GMSYAKGHPSNAFGRATAE--GTYSLAVGLTATAKASSIAGVSNQAIQFATAV 257
Db 1285 YGATVYLDGWSVGPADVTNMPAGTVNIAVSGTNTAGTTISI--SHVYVDLAAVAI 1341
QY 258 -----GGSTQVNLNRGIALFGSQ-----VLQRN---DVNA 286
Db 1342 TINTLSTDVYINAEKSDQLSGTSGVEAGQITVIFPGKSVTTVADNFTWGLTIPA 1401
QY 267 ANPRAVAPDDNPIDNKTKATFKKAGADV-FSIGNSGNDSTIRKLIIVAGS-----338
Db 1402 VDV-----ATLPDGAANVOASVSNVAGNSTOATHAVSDATPASYTINT 1445
QY 339 -ADTDVAVNAQLKEAVLANQ-----ITFKGDSNNRVEKLGKTLITGG-----384
Db 1446 IATDDIINAAAGSALITSGTAEACQIVYTLINGVYNGVQADGMSVSVPTGLAS 1505
QY 385 -----AOTSALTDHINIGVONGDGLKQALLETLSLKMVTENL-----423
Db 1506 LTASSTVYNASVDKAKNSASATHN-----LTVDLAAPVYINTVAGDDIINATEH 1556
QY 424 -----TANE-KVTYKTRLTLDKIGFTNDMNGIDESKPYLDKDTGIIHAGQ 468
Db 1557 GQAQIISGATGATTCMTVSTVITGTYTT-----VIDANG-----NMSIGVPAS 1601
QY 469 KITKLTAG-----VYDDDATYQOLKKVQTAES-----ALQFTYKVKYKNGNDANDSK 518
Db 1602 VISALAGDVTITATVSDAGNSGTASHYTVVALGAPVLAINTIAYDDIINAEKADLA 1661
QY 519 IITVGRNNKPGQVNTLKLKENGVDVTTETNG--TVTFGLNONGNL-----TVGNSTL 571
Db 1662 I--TGTSNQPAQTOI-ITVTLNGON-YTTTADASGMSVYTPASVSLAGEATVYTAAT 1717
QY 572 NNDELIVKNTNSNKOIOVGADGTF---TD-----ISKSPQAGIENTTRI 614
Db 1718 PADGNS-GASHNVQVNTALPGVITINVATDDIINAAEAGVEQITISGQVGAAGDITVY 1776
QY 615 TRDGIFRANNKSIDAKKRLTPGGINA-----GKELTNVQSANPAPFNG-----661
Db 1777 F--LGGTATYVAQANLSVVDYPASALDELGELTISASVYNSVGNNGNGTRETID 1833
QY 662 -----QIDFNNRLSTANTEKSGSAATI---KQLYMLSGVPLTFAGDT-----700

```

RESULT 9
 B85547
 probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substrain
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: B85547

Db 1834 ANLPGLRVDTAGDDVYNIIEHQALVITGSSSLGAGSNVLTINGQTVAAVLADGTW 1893
 QY 701 ---GPNVTKK---LGEILWKKGGKTTAD-----DLTKNNI---GVYADSTDNLSL 740
 Db 1894 SVGVPAVDYSAMPAGSVTTAASGTSAGNPVSYTHEVYVDLSAVASINAITADVDYINA 1953
 QY 741 TVKLAKTSLDLDVANTKTLTASDKVTVDSGNNT-----AKLQND 780
 Db 1954 EKGALTLTSG---STSGVAGQTVTVFGKTSATVAANGSMSTSVPAAMALROD 2009
 QY 781 LT---EKKQNTGATPATNSTGTVD-----GLKTTDNN--- 810
 Db 2010 ASAQASVSVNNGNSATTHAYSDASPTVITNTIAGDILINAEGAALLTIGSSTAE 2069
 QY 811 ---GIALDGTTYITKDKVGFAPKODSLDKSPYLDKDL-----KVG--- 849
 Db 2070 GQVYVTLANGTN---IGYQTDGSMSSVSPSADLSLTLSXYTVANAASVDKAGNPAS 2124
 QY 850 -----EVEITTINGINAGGKAITGLSNTLT---DATNATGHVYQLGIVDST--- 892
 Db 2125 VHNHNLTVDTSVPVYVINTVAGDDVINATEHQAOIISGSAATGATGTVTIGTNTFTT 2184
 QY 893 ---DKTAAASIG-----DVNAGEFNKNDADPEVS-----TYD 924
 Db 2185 VLDASGMSVGPASVYALANGVTYINASVTDAG---GNSGSAHQVYVNTGLPTIIFN 2241
 QY 925 TV---DFINGNATTAKTAVTDGKASKVAYDVNVDGTTIHLTG-----ADGNKNGIGV 973
 Db 2242 AISGNDIILADEKGOPLTISGSGTGLATGAY---TVILINGINYSATTDASN-----W 2292
 QY 974 TTTTLLTKDAKDKAINEFSVNSGDDKALINAKDIADNL-----NTLAGE---IRNT 1020
 Db 2293 TLTVPVSDLAALGOANYTYVASATSAAGNTASSQANLWDSGLPDTINTVAGDDIINA 2352
 QY 1021 KGTADALOTPOVKKVKEGDDNDADTTVGKDAKTNGVNT-----LKLK 1068
 Db 2353 EKGADDTISGVVTRAA-----GDTVYTLGNTYATVQSNLSWSVSPVPTADQALGN 2406
 QY 1069 NGDLID---TNKDGVTVEGINTQSLKAGNNTLNNN---GLSIRKTAGNEQIO----- 1116
 Db 2407 GDLTITASTYTNANG-----NTGSGTR---DITIDANLPLGLAVDVAGDDIYNLSIEHQA 2457
 QY 1117 -----VGADGVKFAKVNNGVAGI---DGTTRI-----TRDEIGFA 1150
 Db 2458 LVITGSSGILNAGAVLTVINSVAATQVADGSMSSVGPANVAMPAGPLTVEVDQS 2517
 QY 1151 GNGSLDKSKPH-----LSKDGINAGKKTNTQSGEIQ---NSNDAYT--- 1192
 Db 2518 SANMPVSVSHPTVDLTAVASINTVASDDVINAEKGTNLTLSGTSIGSGQTVYTF 2577
 QY 1193 GGRKY-----DLKT-----ELENKISSTAKTQNSLHEFSVADEQGN 1229
 Db 2578 GGRKTYASVAANGSMVNPADLTLPEGANVQASVSGNSASATHASV---DANP 2636
 QY 1230 NTFVSNPSSYD---TSKTSDVITTAG---ENGITTVYN---KGVVRVGIQDTKG---LTT 1278
 Db 2637 TLTIT-NTIASDDILNAAEAGSPITISGTASTAGTQTVTLTGATYGTQVADGSMSSV 2695
 QY 1279 PRLTVGNNNGKGIIVD---SONGONTIIGLSNTLANVNDKSVTTEQGIKIKEDTR 1335
 Db 2696 PYSALGALNASYTVSATVNDKAGNPGSHNLAVDTTAPVLTITVAGADDITINAEHQ 2755
 QY 1336 AASIVDVLSAGFNLOGNEAVDEVS-----TYDFVNFADGNATTAKTATYDTSKTSKV 1389
 Db 2756 AL-VISGTSAG---GEXGDVSVVLNGKTYITTLASGMS----- 2792
 QY 1390 YVNVNDOTTIEVKDKKLGKKTTLTSTGTA-----NKFALSNQATGALVKASDIV 1441
 Db 2793 -----VGVPADVYALGSGAOTITASVSDRAGNSDDASRTVYVLSAPV 2836
 QY 1442 AHLNLTSGD---IOTAKGASQANSAGVVDAGNKVYIDSTDNKYYQAKNGQNDVKTREV 1498
 Db 2837 ISINTIAGDVINATEKSDIALS---GTSDDPAGTAITVTLNGQNTSATTDASGMSVTV 2894

QY 1499 AKDKLVAQAPDGLAQMNKSVINKEQVNDANKKGINEDNAFVKLEKASDNKTKN 1558
 Db 2895 P-----ASAVSALGEPATYSVTASVTN-----AAGNSSTASHNOVN 2930
 QY 1559 AAVTVGDLNAAVQAPLFPAGDTGTAKKLGTELTIKGGQTDNKLNTNNIGVAGTGGFT 1618
 Db 2931 TALPEITINPAVTDIINASAGSAQITISGQVTAAGSTYVEL-----GKTYT 2981
 QY 1619 VKLAKDLTINLSVNAAGTKIDDKGVSEYDSSGQAKANTPVLSANGDLGKRYISNNGKT 1678
 Db 2982 ATVOADLSWNVSVPAADMO-----ALNGEELTVNASVTNA-----VGNTEGCT 3024
 QY 1679 KD---TDAAVNOOLNEVRRLGL---GNAGNDNADQONVNIAD-IKKDPNGSSSNRTYIK 1732
 Db 3025 RDTITDA-----SLPELRDVTYAGDD-----VYNIIEHQAOVITGSSSG---FA 3066
 QY 1733 AGTVLGGKGNNDT---EKLATGVGVGVKDKGNANGDLSNWWYTKODGSKKALLATYNA 1789
 Db 3067 AGTALTVYINNQTAAATVLANGSMSSVGPAT---DVSN-W---PAGTLNITYSGANS 3116
 QY 1790 AGQTYNLTN-----NPAEALDRIN---EGGIRF-----FHVNDGQDEPVYOG 1828
 Db 3117 AGTQTSIHPLTVDLTAVAISMNSITSDDAINAEKGAULTLSGTSVGEAGQTVYVTFG 3176
 QY 1829 RNGIDS---SASGKHSY---AIGFOAKADGEAAVAGHOTOAGNQSIAIGNAOTGDOSI 1883
 Db 3177 GKTYTYYAANGSMSTVYPADLALRDGDASAOV-RVTYVN-----GNSATRIHETYSV 3229
 QY 1884 -----AIGTVNVVAKHSGA---IGDPSTVKADNSYSV---GNNQPTDQOTDV 1927
 Db 3230 DSAAPVTYINTIASDNIINASEAAAGVTASGTSAGQGLTLYTLNCTNYQT---VQTD- 3286
 QY 1928 FGVGN-NITVYESVVALGNSAISACT---HAGTQAKKSDG-TAGTT-----TTAG- 1974
 Db 3287 ---GSMSTLTPASDLTALANGTYLTFTVSDLAENLGASAKGVTDTTAPVISFNTVAGD 3343
 QY 1975 -----ATGVKNG-----FAGQ-----TAGVAVSAGASAEIRIONV 2005
 Db 3344 DVINNVHEIQAOIISGATGATGAVAGDRLVYITAGQYVTSIDASGMSVGPASV---ISGL 3401
 QY 2006 AAGEV---SATSTDAVNGS-----QLYKATQGIANTNELDRIHONENKANAGISSAM 2056
 Db 3402 ADGTVTISATITDAGNSNSQTHNVQVNTAAVSLSVTISGDNL-----NAAAG--SAL 3455
 QY 2057 AMASMPQAYIRGSMVMTGAT---HNGGAVAVGLSLNSQGM 2098
 Db 3456 TLSC-----TGTFATGTVYVTLNKGYSAT---IQSGSN 3489

RESULT 10

A41477
 190K surface antigen precursor - Rickettsia rickettsii
 C:Species: Rickettsia rickettsii
 C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 08-Oct-1999
 C:Accession: A41477
 R:Anderson, B.E.; McDonald, G.A.; Jones, D.C.; Regnery, R.L.
 Infect. Immun. 58, 2760-2769, 1990
 A:Title: A protective protein antigen of Rickettsia rickettsii has tandemly repeated,
 A:Reference number: A41477; M01D:90354033
 A:Accession: A41477
 A:Molecule type: DNA
 A:Residues: 1-2249 <AND>
 A:Cross-references: GB:M31227; NID:g152465; PIDN:AAA6380.1; PID:g152466
 A:Note: the authors translated the codon GAT for residue 430 as Gly, and CAA for resi
 C:Keywords: surface; antigen; tandem repeat
 F:1-20/Domain: signal sequence (uncleaved) #status predicted <SIGU>

Query Match 5.9%; Score 630.5; DB 2; Length 2249;
 Best Local Similarity 23.2%; Pred. No. 1,2e-16;
 Matches 500; Conservative 212; Mismatches 779; Indels 665; Gaps 109;

| | | |
|----|------|---|
| OY | 180 | LADINLOKRYRTAOG-----HASTAV-----GAMSYAG-----HFSNAFOTR- 218 |
| Db | 1 | MANISPRLEFKAIQOGLKALFTTSTAALMSSSALGAVATLTANNAAFSNVGNNN 60 |
| OY | 219 | -----ATAEGT-----YSLAVGLTPTAKAASSIIVGNSQAQIGPAAAVGSGT 261 |
| Db | 61 | WNETIAGVANGRPAGPQNNNAFTYGGDYTYTADAADRIIKALN-----VAGTT 110 |
| OY | 262 | OVNIN-----RG-----IALGESSOVLOKONDVNAANRARAAPDDNDPIIDNRK 305 |
| Db | 111 | PVGINTIQNTVGSIIITKGLMLPVTILNAGSKLTLGNNAVNAHGFADADNY----- 163 |
| OY | 306 | ATFNKATDVESIGNSNGDSIRKTIIVNGASADTDVANAQJKEAVRLANROITEKGD 365 |
| Db | 164 | -----TGLNIALGANAAL-----TIGSAAPSKTLIGNID-----GGIITLYKTD 205 |
| OY | 366 | DSNNRVEKGLKTLITITGGAQTSALTDHNIIVGONGDGLKYOLAE--TLTSLMVTY- 420 |
| Db | 206 | AAINGTIGNTNALATYVAVGATITLG---GAVIKATITTKLTNMAASVLTITNANAIVLGA 261 |
| OY | 421 | -ENITLANEKVY---GKTRLTPTDKIGFTDNDNGIDESKPYLDKDTGIHAGQOKITKLA 475 |
| Db | 262 | IDNTTGGDNGVLTNLGALSQYTGDIGNTSLATISVIG-----AGTATLGCAVIKATY 314 |
| OY | 476 | GVYDDDAATYGOGLKRYNQ-----ASALQITTYKVKVDKNGDADSKIIIV-----G 523 |
| Db | 315 | TKLIDAAASAVKFTNPVYVYGCALIDNTGNANNGIYTFEGNST--VTGVNGVTNALATYVAVG 373 |
| OY | 524 | KNNKPDG--TOVNTLKLKGENGVDYTFETNGYTFGL-----NONGLT--VGNSTLUNDG 575 |
| Db | 374 | LLOYOGGVYKANTINL--IDNASAVTFTNPVYVYGCALIDNTGNANNGIYTFEGNSTVIGD- 430 |
| OY | 576 | LSVKNNTSNKOIOVG-----ADGIFPTDISNSKPGAGIENTTRI 614 |
| Db | 421 | -IGNTNALATVYVAGATATLGCAVIKATITTKLTNAAASVLTITNANAIVLGA-IDNTT-- 485 |
| OY | 615 | TRODGIERNNTGSLIDANKRPITPTGINAGKELTNVQASINPATNGGOLDPMNRILSTANT 674 |
| Db | 486 | GGDVGAVGLNTLNGALS-----OYTGNIIGNTSLATISVAGATATLGAV-----IKATTT 534 |
| OY | 675 | EKSASATIKOLYMLSOVPLTFADGPDPVNTKLGIEILVKRGKTTADTLKRN---IG 730 |
| Db | 535 | KLTDAASAVK---FTNPVYVYTGALIDMTGAN--KCIYVTFGNSYVTEGDIGNTSLATIS 588 |
| OY | 731 | VVADSTNSLTVLAKTSLDLDAVNTKTLTASDKV--TVDS--GNNTAKLO--NGDITF 783 |
| Db | 589 | VGAGATILGCAVIKATITTKLTNAAASVLTITNANAIVLGAIDNTGGDNGVLTNLGALS- 647 |
| OY | 784 | SKONTGATPATNS-KTIGCVGLKFTDNNNGIALDGTYYITKDKVYGRAKODGSLDSKPYLD 842 |
| Db | 648 | --OYTGDIGNTSLATISV--GAGTATLGCAVIKATITTKLTNVAASV-----FT 693 |
| OY | 843 | KDKLKGVEVEITTINGIN-----AGKALITG-LSNTITLDTAT-NATTGHYTOIGIY---DST 892 |
| Db | 694 | NPVYVYTGALIDSTGANNGIYTFEGNSTVYIGDIGNTALATYVAVGATATLGCAVIKATIT 753 |
| OY | 893 | DKTRAASIGDVLNAGFNILKNNGDAKDQFVSTYTDVEINGNATTAARVYD-GKASKVAYDV 951 |
| Db | 754 | KLTNMAASVLTITNANAIVLGT---AIDNTGGDNGVLTNLGALSQYTGDIGNTSLA-TI 809 |
| OY | 952 | NVDGTTIHLGADGNKNOIGYKTTTLTKTIDAKDKRKAINESVNSGDDKALINAKDIADNLN 1011 |
| Db | 810 | SVGAGTATLGCA---VIKATITTKLT-----NMAASVLTITNANAV----- 845 |
| OY | 1012 | TLGAIIRTKCTADTALOTFOYKVKVENGD--DDNDADITFYGDKAKT-----NOVNTLK 1064 |
| Db | 846 | -LTGAVDNTTGGDNGVLTNLGALSQYTGDIGNTSLATISVAGATATLGCAVIKATITTK 904 |
| OY | 1065 | LKKGNGIDIONTKDGTVFGINTOSGLKAGNNT-TLNNGL-----SIKNTAGNEOIOY 1117 |
| Db | 905 | LTNAAASVLTITNANAIVLGAIDNTTG---GDNGVLTNLGALSQYTGDIGNTSLATISV 961 |
| OY | 1118 | GA-----DYKFAKVNNGVYVAG-IDGTTIRITRDELGFAGCTN---GS 1155 |

| | | | |
|----|------|---|------|
| Dh | 962 | GAGATLGAIVIAATPTKTLDAASAVKFTNPVYVGAIDNIGNANNIGVTFEGNSVTG | 1021 |
| Qy | 1156 | LDMSKPLSKDGINAGCKRTTNQSEIAQNSNDVATGKTYDLKTELENNISTATAQ | 1215 |
| Dh | 1022 | VGNNTN-LAAVNVAG--LLQYQGVVAKNTINLTD-----NASAVFT | 1062 |
| Qy | 1216 | NSLHEFVADQGNNTVSNPYSVYTSKTSQVTFEAGENGIT-----TVNKGW | 1266 |
| Dh | 1063 | NPVVYVTAIDNTGN-----ANNGIVTFEGNSVYTGAVGNTALATVYNG-- | 1106 |
| Qy | 1267 | RVGIDQFKG-----LTPPLTVG-----NNNGKGIVIDSONGONTIT | 1303 |
| Dh | 1107 | #AGLLQYQGVVAKNTINLTDNASAVTFNPVVYVGAIDNIGNANNIGV--TFNGNSTV | 1163 |
| Qy | 1304 | G-LSNTLANTNBKGSVYRTTEQCKIIRDEKTPRAASYD-----VLSAGFNLQNGEAYD | 1357 |
| Dh | 1164 | GDIGNALATVYVAGITLQAGSL-----AANNIDFGARSTLEFNGPLDGGKRAIP | 1216 |
| Qy | 1358 | F-----VSTYDVYNEADGNATTKAVYDDTSKTSKVY | 1390 |
| Dh | 1217 | YEFKGAIANGNMILAVNTKLLASHLTGTVAEINIGAGNFTTIDASVGD-----VT | 1259 |
| Qy | 1391 | DVANDDTTIEVKQKCGKCTTTTSTGTGANKFALSNQATGALVAKASDIYA----- | 1442 |
| Dh | 1270 | ILNQNINFRARDSVLYSLM-----TGVYNNILL-----ADLVAAPGADETV | 1314 |
| Qy | 1443 | ---HLNTLSGDIQTAKGASQANSSAGVYDADGNK---VIYDS---TDNRYQAKNDGT | 1491 |
| Dh | 1315 | VFNGGVVGLN-----VGSVAVGTARNIGGGCKNFNTLLIYNAVITIDVNLLEGIQWL | 1368 |
| Qy | 1452 | VDKTEVAK-----DKLVAQCPDSTL---AQMNKSVYINKEQVDA | 1531 |
| Dh | 1369 | INKNADPTSTPAENAGAIQIDNATPYTIDANNGNMLNPAENIDFAHADQOLVQNSSGDR | 1428 |
| Qy | 1532 | NKKQGINEDANFVKGLEKASDCKTKNAATVGDLENAV-AQTPLEFAGDGT---AKKL | 1587 |
| Dh | 1429 | TITLGANID-----PDNDDEGIY-----LMSVYAGKKLIIAG--KTEGGAHKL | 1471 |
| Qy | 1588 | GETTLIKG-QGDTNKLITDNNIGNVAGTGBTVYKLAKDLJNLNLSVYAGGTYIDKGVSEV | 1646 |
| Dh | 1472 | -QTILFAGAGDCST-----AGTFENFTNIVLIT----- | 1499 |
| Qy | 1647 | DSSQAKANPTVLSANCLDGKVIYSWGKGTCDTDAVVOOLNEVRNLGLGNNG---N | 1703 |
| Dh | 1500 | --GQ-----LEIG-----ATTAVVLENDAVOLTTQGNIGGFLD | 1531 |
| Qy | 1704 | DNAD-----GNOVNADIKDPNCGSSNRKRYIKAGTVLGGKGNNDIEKLA---TGQVQ | 1754 |
| Dh | 1532 | FNAKNGVYTLNNNVYVAGAVQ--NTGTGNTGTLI-----VLGASNLNRVGIAMLKVYAGN | 1585 |
| Qy | 1755 | VGVDRKDGNAV-GDLSNVYVKTQKDGSKKALLATYVNAAGOTNYLTJNPAEADIRINEGIR | 1813 |
| Dh | 1586 | VTIAKKGKVKIGELQGVGTMT-----LTLPAHFNLTG-----SINKTGQALK | 1628 |
| Qy | 1814 | FFHYVNDQNOEYVVOGRNGIDSSASGKHSVAIGFQAKADEBAVALGRQTOAGNOSTAIGD | 1873 |
| Dh | 1629 | LNFMNGGVSQV-----GTANSVGDIITTPAATISFASV | 1663 |
| Qy | 1874 | NAOAT-----GPOSTA-----IGGNVYVAGKHGALGDPSTYKADNSVSVGNNOFTATQTD | 1926 |
| Dh | 1664 | NKKGTAITLGGTSPANTFTNTGAVTLAK-----GSITSFANVTAT | 1704 |
| Qy | 1927 | VFCVGNNTIVESVVALGSNSAISACTAGTQAKKSDGTAGTTTAYAG---TGTVKG | 1981 |
| Dh | 1705 | SP-VANASNTIFNSLSLAFNSR--ITGS-----GFTLLIGANQVYTYGT--GS | 1746 |
| Qy | 1982 | FAGOTVANGAVSVA--SGAERRIQNNVAAAGSVASTSDAYANGSOLIKATQGIAMNAN | 2035 |
| Dh | 1747 | FTDILLNTFTPDGAAGSGGNILIKSGSTILDLSGVSTLAL-----VVTAN | 1791 |


```

Db 1297 L-----LTHTVAG-----TSNVATIDTISANIDTA 1322
QY 743 KLAKTLSDDAVATKTLTASDKVTYDSSGNNTAKLO-----NGD-----LTFESQNTGA 790
Db 1323 FVA-----GAATATLTPAVNGAAGADGTNOVDALVEDANGNPITGAAYVFSSAN-CA 1375
QY 791 TPAATSKTIGVDLKEFT-----DNNGIALDGTYYITKDKVGFAGKODGSLDKSKPYLD 842
Db 1376 TILSTMTNTGVASTFLTHVAGTSNVVATIGSTEENIDTAFAGANATITLAPV-- 1433
QY 843 KOKLAKGEV--ITTINGIA-----GKAITGLSTLNDATNATGHTVQLGIVDSTDKT 895
Db 1434 -----NGAAGDGVNTNSVQAVVSDSDGNNAVGTATVFSSA-NATAQITTVIG-ITGAGGI 1486
QY 896 RAASIDVYNAGFNKLNNGDAKDEFTVDYFINGNATATKATVTDGKASKAYVNDG 955
Db 1487 AATRLTINIVAGTSNV-----VATIDTV--NANIDTTPVA--GELENTIV--VSTIN 1530
QY 956 TTHLTGADGNKNOIGKTTTLTKTDKAKDAINFVSNGDDKA-LINAKDIADNLNTLA 1014
Db 1531 NNALAGADPTNIVEAFV-----TDRENGVANOGLIFGTNGASIVGSSVTITLND-- 1580
QY 1015 GEIR-----NTGCTADTAL-----GTQVKKYK-----ENGDDND 1045
Db 1581 GRVRSATHTVAGSSNTVIAISGAGYARVTFVADVSTAQLKLTSLDNOJLANKAGNI 1640
QY 1046 ADPTITVGRKAKTNOV--NTLKIKKKNGLDIOTNKDGTVPFGT-----NTQSGT----- 1091
Db 1641 AQAALVY--DAHNLNLANOSVSFALDNGAVIESOGDSASSAGIVLAKRPNNTLAKMTVTAT 1698
QY 1092 --KAGNNTLLNNGT-----SIKNTAGNEQIOVGADGVYFAKAVNNGVVG- 1133
Db 1699 LDSTGQTELETHFPAVKAASITEMTWMKNAYANNIDTNEVOVLVTDGNNAINAVNL 1758
QY 1134 -----ACIDG--TTRITRDEITGAGTNGSLDKSKPHLSKDGINA---GCKK 1174
Db 1759 TNSGNNITPNSVTGSDGTATATLHTLAGSLPINARDQ---VSKT-INATFIADAS 1813
QY 1175 ITNIOGSEIAONSNDPVTGKITYD-----LKTLENKISSAKTAKOASLHEF 1221
Db 1814 TQOILAGDFTIYNDQVANGQAVNAVQARVNDSTGNPIKDDTVEVFLSNGITQELDVT 1873
QY 1222 SVADGQNNFTVSNPXS-----SYDTSKTSDDVIT-----FAGENG 1256
Db 1874 SV--EGVAVTEFTNLTAGITNTATVAVSSGSSRNIDTFTFADVTAHLAASDLAMIVDA 1931
QY 1257 ITTKVAKGVYRVIDOTKGLITPKLTGVGNNGKGIIVDSQNGQNTITGLSTLANVTMDK 1316
Db 1932 VADNLDKNEVHARVTDKGNVLSGQVIFTSNGGAITTVNGISDGDGLTKATLHTLAG 1991
QY 1317 GSVRTTEGKIITKDEDKTFAA--STVDYLSAGFNLOGNEAVDFSTYDT--VNFADGN 1371
Db 1992 TSVVAVRAGNRQOSKDTFTIADRTATIRASDLITTRNNALADGVAITNAARYIVTDAGN 2051
QY 1372 AFTAVVYDDTSKTSKVYVDVNDVDTTEVKDKKLGVTTLTSTGTANK-----FALS 1426
Db 2052 PVPSEVFGTSDNGALIFPTSGMTDS-----GTSTFTHTTAGISCTAIVTMG 2103
QY 1427 NGATDADLVKADIVAHNTLSGDIOTAKGASQANS-----AGYVADGN-----KYI 1475
Db 2104 ISQTDAAFIARSTAHVS---ELIVVKNSTLANSRNIYQAHIKAHGCVVVTGMVNV 2159
QY 1476 YDSTNNKYVQAKNDGTVDKTEKVEADKILVAQOTPDGTGLAQMNVSVINKEDVNDANKQ 1535
Db 2160 FSATENVTILTANTVT--NSQGYAENTLRHNAPTVSAITAIV-----ATLIV 2204
QY 1536 GINEDNAEYKGL-----EKAASD-----NKTNAAVTVGDLNAAVQPLTFAGDTG 1581
Db 2205 GLETEVREVAAGARIELEFRINDGAVADGICITNREARVYDSD-NLVPNNSVVFESADG 2263
QY 1582 -----TTAKLIGET-LTIKGGQDITKLDNNTGVAGT-----DGFYVTKAKD--- 1624

```

```

Db 2264 GOLVONDVQDALSAAVTVSNINTGVTKVTADGVASATTTFTIADRDATLVTRFL 2323
QY 1625 LTN-----LNSVNAAGTKIDDKGVSEFVSSGAKNTEVYLANGLDGKV 1670
Db 2324 ITHDNAANGVENRVLHLVDADNSVSGVEVFSAITNGASINASATIDINGPAIG--V 2381
QY 1671 ISNVKGTGKTDDANVOOLNEVRNLGLGNAGNN-----ADGNQVNIADIKKDPNS 1722
Db 2382 LNTLSGSDVTVLVLP-----CGTESLVTTPQFIADINTANIA----- 2421
QY 1723 GSSNRYVYTKGYLVGKGNNDTEKLTAGVGVGVKDNAGNDSLAVWK----- 1773
Db 2422 --TGDFVIIDGAVANSYDANEVARYT-----DMQNALIAGSYVFESSONGATITTS 2472
QY 1774 --TQKDSKRALATPYNAGOTNTLN-----NPAE 1802
Db 2473 GITVDGKASAKL--THIKAGESGLALSRPMATVHTLMPFIADSTATLOLFENRP-- 2529
QY 1803 AIDRINEGIRFE--HYNDGNDEPVYVQGRNGIDSSASGRHSAVIGFOAKADEBAVAIG 1859
Db 2530 -IPIIADGVMOFFLGRVFPANQNPV--GGQVAFSATNEVITLESNGISTEGSVLLS 2586
QY 1860 -KOTQAGNOSTAIDNQAOTGDOSIAGTGVYVAKHSGAIG-----DPSTVKAADNSYV 1913
Db 2587 VTSTQAGVHPF-----TGLVSNNTYDTPFGAFTANKNTAOLTIMVY 2629
QY 1914 GNNQFDPATQTFVGVGNNTTVESNSVALGSASISACTHGTQAK-----SD----- 1964
Db 2630 DNMLADCVTRNOVRA---HYVDSTGNSVA---DMAVTFANRAQLSKTVYTLDDNGDA 2683
QY 1965 -----GTAGTTTGTAGATGVYKGFAGOTAVAGV--SVGASGERRIQ- 2003
Db 2684 VNTLTNSLVGTVYATLGTAGPFLVDYFTAGPLATLTLVTVNNAFADNSATNTVQA 2743
QY 2004 -----NVAAGEVSAISTDVAVNSQLYKATQGIANATNELDRIHONENKANAGISSAM 2056
Db 2744 TLKDVSGNPVIGEVAFRA--ASNATITTAARDGVSNAG--IYATLTNGTAGSVTV 2797
QY 2057 A-----MASMPQAYIIPGRSM-VTGGIATFHNGQGAVA---VGLSKLSDNGQWFKINGSA 2106
Db 2798 ATTELTETTTDTTIANKNLDVYVNTGTTFGNDAGFPPTGTVGAT-----FKVNSGG 2848
QY 2107 D 2107
Db 2849 D 2849

```

RESULT 13

```

F90073
hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F90073
R:Kuroda, M.; Ohca, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani, U.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; PMID:21311952; PMID:11418146
A:Accession: F90073
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2271 <KUR>
A:Cross-references: GB:BA000018; PID:g13702612; PIDN:BA843752.1; GSPDB:GN00149
C:Experimental source: strain N315
C:Genetics:
A:Gene: SA2447

```

```

Query Match          5.5%; Score 590.5; DB 2; Length 2271;
Best Local Similarity 17.0%; Pred. No. 4,1e-15;
Matches 385; Conservative 435; Mismatches 973; Indels 469; Gaps 78;

```


Db 2146 STSMGSGSISDSTSTSMGSGSTSTSESNMHPDSMSMHHTSTSTRLSEATTSTSES 2205
 QY 2063 GAYIPGRSMVTGGIATHNQGAIVAGLSKLSNNGMWKING 2104
 Db 2206 QSTLSATSEVT---KHNGTPOASE--KRLPDTGDI--KONG 2240

RESULT 14 AB3528

extracellular serine proteinase (EC 3.4.21.-) [Imported] - Brucella melitensis (strain 1
 C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 C:Accession: AB3528
 R:DelVecchio, V.G.; Kaprat, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova,
 R.; DelVecchio, V.G.; Goldsman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
 A:Reference number: AD3252; PMID:11756688
 A:Accession: AB3528
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2554 <KUR>
 A:Cross-reference: GB:AE008918; PIDN:AA153389.1; PID:917984282; GSFDB:GN00191
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BMEI10148
 A:Map position: 11
 C:Keywords: hydrolase; serine proteinase

Query Match 5.3%; Score 568.5; DB 2; Length 2554;
 Best Local Similarity 21.4%; Pred. No. 3.4e-14;
 Matches 492; Conservative 271; Mismatches 925; Indels 607; Gaps 104;

QY 202 GAMSATKGFNSAFGRATAECTISLAVGLTATKAASSIANGSAQAIGFAAT- 256
 Db 20 GAIGFTR---AGAGSVTLTGQNTYSGVTLTSGILTVTOMANAGASGIG--OSTADPAWL 75
 QY 257 ---VGSSTOVNLNRGIALGFGSOVLQKNDVNAANRAVAPDNOPIDNRYK-- 305
 Db 76 MLESCTFRITGGS--VTTRGFTLVNGPARIETVGTSSNL--AFSGLVTSDDAGFEK 132
 QY 306 ---ATFKNGATDVESIGNSGNDSIRKRIIVG---AGSADTAVVNAQLKEAVRL-- 355
 Db 133 GAGTTLFLNGSNDHIGATVSGTILAVSTLADGQVSSLGKSGSDATNLLIAGALNTYG 192
 QY 356 ---ARQITFRKDDSNRNVEKGLKTLITGGAQTSALTDHINICVQGDLLKQLAET 411
 Db 193 STTSSRSFTLGAAGSISGVANA--GTTLSMSG--TAVGTD--GLTKLGDG----- 237
 QY 412 LNSLKAVTENTLANEKVTVGKTRLTDDKI---GFTNDMNGIDBSKPYLDDKGTIHAGG 467
 Db 238 --TLISGTNTYTGNTAVAGVLRAGSAQAFPSGLMTVNG-----ASLELGG 284
 QY 468 QKIT---KLPAVDDDAATYGOLKKVNOFASALOTFTVKVVDKNNDANDSKITTYGK 524
 Db 285 YDITVSGLLGAGVLDGNTL-----TSSGSAANSFTGKITGGFTRTGSGTQTLISG 337
 QY 525 NKRPDQVNTLKLKENGENDVDTTETNGVTVGLNQN--NGLTVNGSTLNNGLSKNN 582
 Db 338 CNS-DYTGKTTTASNGTSLVDCLKNGQASSIGASSNADPNLVNNGTSLTYGNTV--TTD 395
 QY 583 SNKOIOVAGDGIPTDISKSPKAGIEGTRITRDIGIPANNVTSL--DANKPRLTPG 639
 Db 396 RGTIDGGTCAISVTAATTLTFSG-----QYVGTALOKRGTGLVLMNSYSR--GGTS 449
 QY 640 INAGKELTNVOSALNPATNGOLDPMNRLSTANTEKSSAATIKDL--YNLSQVPLFA 697
 Db 450 VDAQ-----TLRAGSSGARGGSGMSLSN-----AAGAILDLDFDFTSVTSLISG 493
 QY 698 GDTGPVNTKRLGILKVGKGTADLT-----KNNIGV-----VAOSTNSLTV 742
 Db 494 GALGGVVALG--GATTLTISGNSNGTSTYTGATIGNFVKNGTORLTGCASSYSQSTTI 552

QY 743 K---LAKTSDHDAVTKTLTASDKVT--VDSGNNTAKLQNGDLTFESKONTGATPATNSK 797
 Db 553 NGVLEVSCLADGGSVSSISMSADADNLVINGVLTLYTGGSDSTROQLTLAGSGNSIE 612
 QY 798 TIVGDLKFTDNNGI---ALDGTIYITKDKVGRAKODGSLDKSKPYLDDKLVGEVEIT 854
 Db 613 SEGTGAILFTSMAAVTTPAANTAGTLTLTLAGTNTDNERG-----AQLT 657
 QY 855 TNGINAGKMTGTSNT-----LTDAFNATGTVHVL--GIYVSTDK---TRAASIGDV 903
 Db 658 NN-----GSGTISLTKIDTGTWFLTNSDSTYTG--VTKINGVLSVDKLVAGVLAASSIGAS 711
 QY 904 LNAENIKNNGDADPEFSTYDVDFINGNATKATKATVYDGKASVAVDVNDGTHLTLGA 963
 Db 712 SSAASNL-----IIGNDSTLRYLGTGDTTDLRLFLTAGSLTYIESGGS 753
 QY 964 DG---NKNQIGVTTTLTLDKAGDKAINEFVNSGDKALINAKDIDLNLAGEIRN 1019
 Db 754 GAIVFTDGOVALADNNAQRTIALGK-----NTGD-----NRLASIGD 793
 QY 1020 EKGADALOTFGYKVKYKENGDDNDADTTIVGKDAKTQNVTLK--LKGNGLDIQTNK 1077
 Db 794 A-GTGTITLA-----KNDGTGVLTGNNITGFTNINKGLKLTGNG-----GT 835
 QY 1078 DGTVTFGI--NTQSLKAGNNTLNNGLSIRKTAGNEQIOVGADGVKFAKVNNGVAG-- 1135
 Db 836 TGSILSDIVATDGLIFNRSDTLNYYGL---ISGAGFVYSGSGFTILGANSYTGATS 891
 QY 1136 -IDGTRITRDEIFAG---TNGSLDKSRPHSKDQINAGGKKTINIGSGEIAONSND 1190
 Db 892 VSAGTLLVNGDSANATQTSVANSI---LEGSGIIGGNVYVD--GALAPSGNG 942
 QY 1191 ---VTGKIKYDKTELEKIS---STAKTAONSL-- 1218
 Db 943 GTLLTNGSLAISASISLMQAGAGVAGALNDL--IEVKNLITDGLVDAEFAAGSYGP 1001
 QY 1219 ---HEFSVADEQGNFTVSNPYSSYDTSKTSVITYPAG-----ENGITRYNKG 1264
 Db 1002 GIYRLINVTGSLTDNGLDIGMLPAGAGIQTAVAGQVNLLAGTNEFMFGDVGKPFNSA 1061
 QY 1265 VVRGIDOTKLTLPKLTIVGNN-----CGKIVIDSONQNT 1301
 Db 1062 V-----DGGNGTWQNSSGNNTWATGATGNINASYSDGAPLFTGATVTDNLSGQVK 1114
 QY 1302 ITGLSNTLAN--VTNDRGSVTRTEQGIIRKDEKTRA-----ASIVDLSAGENLQNGE 1354
 Db 1115 AEGMQFAIDSAVYGDK--LELTGPSTIRVGDSGTGAAYIAVINSVLGNQLE--XTD 1171
 QY 1355 AVDEPSTYDVNFEADG---NATTAKVITYDT--SKTSKVYDVAVNDTTIEVKDKLGVK 1409
 Db 1172 AGTIVLT--GANSYTGALINGTIRISSDNLGYASSDISPDGALMTWTANIAIDRAIIL 1230
 QY 1410 T-----TTLTSGTGANKFALSNOATGDALVYASDVAHLNLTNSGDIOTAKGASQ 1459
 Db 1231 TGAGTLLTDASTTSLSPISGTLKTSGTLL--SGTAH---TGGTITTACTLO 1284
 QY 1460 ANSSAGYVADAGNKV---IYDSTDKYXQAKNDGTVDKTYE-----VAKDKIYAOA 1507
 Db 1265 IGNKGTDSIDGNTVNNCALVFPDRAGLATYGSISGIGTLKKNSSITLWMTGISTYGET 1344
 QY 1508 QTPDGTIAQW--NVKSYIN-----KEQVNDANKKOGINEDNAFVGLKKAASDNKTN 1358
 Db 1345 TVSAGTTLAQAGQIKGASLTVDGAGEVLIDGSGSOFATGAGASVYG-----T 1393
 QY 1559 AAVTYGDLANAVAQPLTFAGGTGTAKKLGTLTIKGGQDYNLTDNNNIGV--VAGT-- 1614
 Db 1394 GTVYVRDGTASFDLSLTNAATGINS---TIVVAG--SGSOMTGTGATFGLAGTAT 1445
 QY 1615 ---DGFTV-----KLAIDLTLNLSVNAAGTK-----IDDKGVS 1644
 Db 1446 VDILDGTMISGASVYRGQDLPMDATGQVYISGAGSOWTIANLVARSGSITYDDGCV 1505

| | | | | |
|---|---|-----|---|-----|
| D | b | 144 | LNQ-----NNADANHGEGAPADNYTGTLNGLGANAALLITQSAAPAKITLAGNIN | 195 |
| Q | y | 116 | NNANIGAKAGNESIAIGC-DVLAASHASIALGSDLLYIAKKTVOQISELLPIIRGOKAL | 174 |
| D | b | 196 | GGGIIYKTKDMAINGITGNATNALATVAVGAGIATLEGAIKATTTKTLNMAAVL----- | 249 |
| Q | y | 175 | NDIYQADMTMOKYRRTTHAGHASTVAVGAMSAKGFHSAFSTR---ATAGEYSLAVG | 230 |
| D | b | 250 | -----TLNNNAVILGTGAIDMTTGVDVAVGLNGLALMSQYTG | 285 |
| Q | y | 231 | LTATAKASSIAVGSNAQAIGFAPATVAGSTOVNLRGIALGFSQVLOK---DNDVAA | 287 |
| D | b | 286 | NIGMTNALATISVAGAKATLGGA--VIAKATTTKLTDMASAVFTPNPVVYTGAIIDMTGN- | 342 |
| Q | y | 288 | NVRAPADDQNPIDNRKATFEKNGAIDVDSIGNSGNSDIRKTIINVGASAD----- | 340 |
| D | b | 343 | -----NNGIYFTGDSYTVTGIGNITNA-----LATISVAGAKATLGGAIIK | 383 |
| Q | y | 341 | -----TDAVNAVQLEAVELA-----NROITFKGDDSN-----RVEKG | 374 |
| D | b | 384 | ATTTKLTDMNASVFTPNPVVYTGAIIDMTGNANNGIYFTGDSYTVGNIGNATLATISVG | 443 |
| Q | y | 375 | LKGLLITTCG---AQTSALTDHNIGVYQNGDLKVLQALATLISLKMV-----PENLTAN | 426 |
| D | b | 444 | AGKA-TLGGAIIRKATTTKLTDMASA-----TFTPNPVVGAIDMTGNAN-N | 488 |
| Q | y | 427 | EKVAVGTARLTTDGIPTGNMNGIDSEKPYLDKDTIHAG---COKITKLAVGVDDA | 482 |
| D | b | 489 | GIYFTGDSYTVTGIGNITNALATI-----SVGAKKATLGGAIIKATTTKLTDMNA | 537 |
| Q | y | 483 | ATYQOLKRVNOT-----AESALOTFT-VKKVDKNGDANDASKIITVYKNNKP-DGTQ | 532 |
| D | b | 538 | SAVFTPNPVVYTGAIIDMTGNANNGIYFTGDSYTVGNIGNATLATISVAGAKATLGAI | 597 |
| Q | y | 533 | VNTLKLGENVDYTTTNGYTFGL-----NONNGLT---VGNSTLNDGSLVYKTFNSNK | 585 |
| D | b | 598 | EKATTTKLTDMASAVFTPNPVVYTGAIIDMTGNANNGIYFTGDSYTVG---NIGMTNALA | 654 |
| Q | y | 586 | QIOVGA-----DG-----ITFEDISNSPGAGIENTYRITRBDGIGFANN | 624 |
| D | b | 655 | TYNVGAGATIEGAVIKATTTKLTNMAAVLTLINVAVILGGA-IDMTGV--DNVGVNL | 711 |
| Q | y | 625 | TGSLD-----ANKPRLTPTGINAGKELINVO\$AT---NPATNGOL | 663 |
| D | b | 712 | NGALSOYTGIGNATNALATISVAGAKATLGGAIIKATTTKLTDMASAVFTPNPVVYTGAI | 771 |
| Q | y | 664 | DEMRRLSTANERKSGSAATIKDLNLSQVLPFAGTGPVWYKILGE-----I | 711 |
| D | b | 772 | D-----NTGNA--NNGIA-----TTGGDS--TVTGNIGNATNALATVAVGAGL | 809 |
| Q | y | 712 | LKVAGKTTAADLTKNNIGVAVADSTNSLTVKLAKTSLDIDAVNFTKTLTASDKVTVDSG- | 770 |
| D | b | 810 | LRYQG-----GVKSWTIN-----LTD---NASAVFTPNPVVYTGAI | 843 |
| Q | y | 771 | NNTAKLONGDITSEKON--TGAIPATHS-KTIGVDGLKFTDNNGIALDSTYITITDKYGF | 827 |
| D | b | 844 | DNTGNANNGIYFTGDSYTVGNIGNATNALATISVAGAKATLGGAIIKATTTKLTDMNA--- | 900 |
| Q | y | 828 | AKODSIDLKSPEYDKRKLKVEYEITTNGINNGKALITGLSTLIDATNATGHVTOLG | 887 |
| D | b | 901 | -----SAVFTPNPV-----VYTGAIIDMTGNANNGIY-----TFTGDSYTVG | 936 |

```

Db 1035 ASAGDVTLINAODI--HFRAL-----DSALVLSNLGTVGN---NILLAADLV 1077
Qy 1050 TVGKDAKTNOVNTLKLKGNGLDIOFNKDGTVFEGINTOSGLKAGN--NTTLNNNSLTK 1107
Db 1078 ABQVDGCTVFD---GGVNGNLISNVAGAA-----KNIGDVGKKEFTLILYNAVTT 1128
Qy 1108 MTAGNEOIOVGADGVFAKYNNGVAGIDGTRIRTRDEIFAGTNGSLDKSKPHLSKDG 1167
Db 1129 D-----DVNLEGIONVLNN---NADFTSSJAFNAGTIOINDATYITIDANNGLN--- 1175
Qy 1168 INAGGKKTINIOSGEIAONS--ND-AVTGKIYDLKTELENKISSTAKIAONSLEHFSVA 1224
Db 1176 IPAGNKKFAHADAQILLONSGNDRTITLGANIDPDNDDEGIILNSVLA----- 1225
Qy 1225 DEOGNFTVSNPYSSYDTSKTSVDITFEAGENGITTVKNKGVRVGDIDQKGLTTPKLTVG 1284
Db 1226 ---GKLLTTAGGKTPGAKRLDIIY-FKGE-----GDFGTAGTTP----- 1261
Qy 1285 NNGKGIVIDSONGONTITG---LSNLTANVTNDKGSVRTTEOGKLIKDEKTRAASTVD 1341
Db 1262 --NTNIVLD-----ITGOLELGATNAVYLFKDAVQLTQTG----- 1296
Qy 1342 VLSAGFNLOGNEAVDFVSTYDTPVNFADGNATTAKYITDITSKTSKVYDVAVDDTTIEV 1401
Db 1297 ---NIGG-----FLDENAKNGTVTLNN---NNVWAGT----- 1322
Qy 1402 KDKKLGKTTTLTSTGT---GANKFALSNQATGDALVAKSDIVAHNLNTLSGDIQTAKGA 1457
Db 1323 ---VKNTGGTNNGTLLVIGASNL---NRVNGIAMLKVG-----AGNVTIAKG- 1363
Qy 1458 SQANSSAGYVDADGNKVIVYDSTDNKYQAOKNDGTVDKTEVAKDKLVAAQOTPDGTLAQM 1517
Db 1364 --GNVKGIGETGTWTL--FLPAHF--KLFGSINKT-----GGQALKL 1401
Qy 1518 NVKSVINKQOVANDANKKOCINEDNAFVKLEKAAASDNKTKNAAYVVDGLNAYQTPLPFA 1577
Db 1402 N-----FMNG-----GSVSGVVGTPANSV 1420
Qy 1578 GDTGTT-AKKLGETLTKG---GOTDTNKLFDNNIGVAGTDFVYKAKDLT----- 1626
Db 1421 GDITTAGATSPASSVANAKGTATLGTTSPAHFTNTGAVTLAKGSITSFAKNTVATSFVA 1480
Qy 1627 NLNSVNAG-----GKIKDKGV-----SEVDS-----SGQAKANTPYL 1659
Db 1481 NSATINFGNSLAFNSNITSGTTLILGANQVYTGTSPTDPLTLNTJFPGAAGSGGNLT 1540
Qy 1660 SANG--LDIGGVVINSVNGCTKDTDAANYQULNEVRNLLGLGNAGDNADGNQVNIADIK 1717
Db 1541 IKSSTLIDLGS--VSNLALVVTAT-----NEDMN----- 1567
Qy 1718 KDPNCGSSSNRTVIRKAGTVLGGK-----NND-----TEKLTAG 1751
Db 1568 ---NISPTKTYVISAETAGGLKPTPKENVKITINDNRFVDFPDASTITLFAEDIAAG 1624
Qy 1752 GVQGVYDKDGNANGDLSNV-----WVKTKDGSKK-----ALLATYNAAGOT 1793
Db 1625 ---VIDEDFAFGGPLANIPNANIKKSLELMEDAPNGSDARQAFNNFGIMTPIQADAT 1680
Qy 1794 NYLTNN---PARAIDRINEOGI-----RFFHYNDNGNEPVYQGRNGIDSSAS 1837
Db 1681 THLMQDVAKPSDTIAVANNQVVASINSNTALNARMKVOAGNKPVSSGDEDMD--- 1736
Qy 1838 GRHSAVIGFOAKADEEAAVAGIROTOAGNOSTAIGDNAOATG-----DQSAIAGTG 1888
Db 1737 ---AKFGAMISPFVGNATQKMONSIS-GYKSDITGCTIGFPGFVSDDLVLGLA 1785
Qy 1889 NVVAGK---HSGAIGDSTYKAD-----NSYSGVNNQFTDA----- 1922
Db 1786 YTRADTDIKLNKKTGDKKKNVESNITSYGLYSPYENLFVEALIASYDNKIRSKSRHY 1845
Qy 1923 ---TQTDVFGVGNNTITVESNSVAL-----GNSAISAGTHAGTQAKKSDGTAGT 1969
Db 1846 ATTLTETVGIQTANGKYYKSSSTYGLMAGYTYMSENINLTPLAGLRYSTIKDKSYKERTG 1905

```

```

Qy 1970 TTTAGATGVKGFAGQTAVGAVSVGASGAEIRIONVAAGEVSATSDAVNSQLYKATQ- 2028
Db 1906 TY---QNLTVKGNKNTNFDGLGAKVS-----SNINVNELVLT-----PELIYANVDY 1949
Qy 2029 GIANATNELDRIHONENKANAGISSAMAMASMPQA 2064
Db 1950 AFKNKVSALDARLO-----GMTAPLPTNSFKRS 1977

```

Search completed: July 30, 2002, 15:37:07
Job time: 332 sec

Wed Jul 31 07:54:12 2002

us-09-813-214a-9.rpr

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:37:59 ; Search time 27.4 Seconds
(without alignments)
2998.645 Million cell updates/sec

Title: US-09-813-214A-9

Perfect score: 10708

Sequence: 1 MHHTYKVFKNKATGTFMVA.....NGSADTQGHGAGVAGRHF 2122

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues 105224

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 630.5 | 5.9 | 2249 | 1 OMPA_RICRI | P15921 rickettsia |
| 2 | 550.5 | 5.1 | 2021 | 1 OMPA_RICRI | O52657 rickettsia |
| 3 | 544 | 5.1 | 2003 | 1 YDBA_ECOLI | P33666 escherichia |
| 4 | 494 | 4.6 | 1656 | 1 OMPB_RICJA | O06653 r outer mem |
| 5 | 493.5 | 4.6 | 1953 | 1 BIGA_SALTY | P25927 salmonella |
| 6 | 480 | 4.5 | 1643 | 1 OMPB_RICRI | O53020 r outer mem |
| 7 | 478.5 | 4.5 | 3591 | 1 FHAB_BORPE | P12255 bordetella |
| 8 | 474.5 | 4.4 | 1577 | 1 HUYA_PROMI | P16466 proteus mir |
| 9 | 456.5 | 4.3 | 1655 | 1 OMPB_RICCN | O9K83 r outer mem |
| 10 | 444.5 | 4.2 | 1902 | 1 P3P_IACIC | P15292 lactococcus |
| 11 | 437.5 | 4.1 | 1902 | 1 P1P_IACIC | P16271 lactococcus |
| 12 | 437.5 | 4.1 | 1902 | 1 P2P_IACIC | O53047 r outer mem |
| 13 | 435.5 | 4.0 | 1654 | 1 OMPB_RICRI | O02470 lactobacilli |
| 14 | 433.5 | 4.0 | 1902 | 1 P2P_IACPA | P69699 r outer mem |
| 15 | 430 | 4.0 | 1645 | 1 OMPB_RICRY | P53320 serratia ma |
| 16 | 425 | 4.0 | 1608 | 1 HUYA_SERMA | P52143 escherichia |
| 17 | 403.5 | 3.8 | 1569 | 1 YPUA_ECOLI | P38058 clostridium |
| 18 | 398 | 3.7 | 1848 | 1 CBPA_CLOCL | O07833 bacillus su |
| 19 | 380.5 | 3.6 | 2334 | 1 WAPA_BACSU | P32051 escherichia |
| 20 | 366 | 3.4 | 1325 | 1 YDEK_ECOLI | P14914 rickettsia |
| 21 | 360 | 3.4 | 1300 | 1 120K_RICRI | P38537 bacillus sp |
| 22 | 354 | 3.3 | 1176 | 1 SLAP_BACSH | P55127 neisseria m |
| 23 | 353 | 3.3 | 1829 | 1 FRPC_NEIMC | O9JYV5 neisseria m |
| 24 | 349.5 | 3.3 | 1829 | 1 FRPC_NEIMB | O92812 chlamydia p |
| 25 | 342 | 3.2 | 1723 | 1 PM20_CHLPN | P35828 caulobacter |
| 26 | 339 | 3.2 | 1025 | 1 SLAP_CAUCR | O03155 escherichia |
| 27 | 338 | 3.2 | 1286 | 1 AIDA_ECOLI | P44969 haemophilus |
| 28 | 329.5 | 3.1 | 1694 | 1 IGAO_HAETN | P18127 xanthomonas |
| 29 | 325.5 | 3.0 | 1567 | 1 ICEN_XANCT | P45384 haemophilus |
| 30 | 325 | 3.0 | 1702 | 1 IGA2_HAETN | P54363 dictyostell |
| 31 | 321.5 | 3.0 | 1849 | 1 IGA4_HAETN | O84419 chlamydia t |
| 32 | 321 | 3.0 | 2492 | 1 TALA_DICDI | |
| 33 | 310.5 | 2.9 | 1770 | 1 PMPC_CHLTR | |

| | | | | | |
|----|-------|-----|------|--------------|--------------------|
| 34 | 309.5 | 2.9 | 918 | 1 YMOB_CAEBL | P34487 caenorhabdi |
| 35 | 301.5 | 2.8 | 1672 | 1 ICEA_PANMU | O9PJY2 chlamydia m |
| 36 | 300.5 | 2.8 | 1322 | 1 ICEA_PANMU | P20469 pantoea a |
| 37 | 299 | 2.8 | 1183 | 1 CNA_STAU | O53654 staphylococ |
| 38 | 298.5 | 2.8 | 1296 | 1 ASAL_ENTPA | P17953 enterococcu |
| 39 | 295.5 | 2.8 | 1256 | 1 MRP_STRSU | P32653 streptococ |
| 40 | 293 | 2.7 | 1228 | 1 SLAP_BACST | P35825 bacillus st |
| 41 | 293 | 2.7 | 1754 | 1 PMPB_CHLTR | O84418 chlamydia t |
| 42 | 291 | 2.7 | 1258 | 1 ICEN_ERWHE | P16239 erwinia her |
| 43 | 290.5 | 2.7 | 1276 | 1 PMPB_CHLPN | O92899 chlamydia p |
| 44 | 290.5 | 2.7 | 1637 | 1 MRSP_STAU | P80544 staphylococ |
| 45 | 290.5 | 2.7 | 3562 | 1 PGCV_CHICK | O90953 gallus gall |

ALIGNMENTS

| RESULT 1 | ID | STANDARD: | PRT: 2249 AA. |
|----------|--|-----------|---------------|
| AC | OMPA_RICRI | | |
| AC | P15921 | | |
| DT | 01-APR-1990 (Rel. 14, Created) | | |
| DT | 01-APR-1990 (Rel. 14, Last sequence update) | | |
| DT | 16-OCT-2001 (Rel. 40, Last annotation update) | | |
| DE | Outer membrane protein A precursor (190 kDa antigen) (Cell surface antigen) (rompa) (romp A). | | |
| GN | OMPA.. | | |
| OS | Rickettsia rickettsii. | | |
| OC | Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; | | |
| OC | Rickettsiaceae; Rickettsiae; Rickettsia. | | |
| OX | NCBI_TaxID=783; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=N. | | |
| RC | MEDLINE=90354033; PubMed=2117568; | | |
| RA | Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.; | | |
| RT | "A protective protein antigen of Rickettsia rickettsii has tandemly repeated, near-identical sequences."; | | |
| RL | Infect. Immun. 58:2760-2769(1990) | | |
| CC | -1- FUNCTION: ELICITS PROTECTIVE IMMUNITY. | | |
| CC | -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A | | |
| CC | -1- LAYER WITH HEXAGONAL SYMMETRY. | | |
| CC | -1- PTM: GLYCOSYLATED (PROBABLY). | | |
| CC | -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY. | | |
| CC | THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION | | |
| CC | BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION | | |
| CC | AT THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS | | |
| CC | USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY | | |
| CC | MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL | | |
| CC | ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/ | | |
| CC | OR SEND AN EMAIL TO license@sib-sib.ch). | | |
| CC | ----- | | |
| DR | EMBL: M31227; AAA26380.1; - | | |
| DR | PIR: A41477; A41477. | | |
| DR | InterPro: IPR003858; rompa.rompb. | | |
| DR | Pfam: PF02708; rompa.rompb.1. | | |
| KW | Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein. | | |
| FT | SIGNAL | 1 | 28 |
| FT | CHAIN | 29 | 2249 |
| FT | DOMAIN | 212 | 1180 |
| FT | REPEAT | 212 | 286 |
| FT | REPEAT | 287 | 358 |
| FT | REPEAT | 359 | 430 |
| FT | REPEAT | 431 | 505 |
| FT | REPEAT | 506 | 577 |
| FT | REPEAT | 578 | 652 |
| FT | REPEAT | 653 | 724 |
| FT | REPEAT | 725 | 799 |
| FT | REPEAT | 800 | 874 |
| FT | REPEAT | 875 | 949 |
| FT | REPEAT | 950 | 1021 |
| FT | REPEAT | 1022 | 1093 |
| FT | REPEAT | | |

FT REPEAT 1094 1165 M (TYPE II).
 FT REPEAT 1166 1180 TYPE I (INCOMPLETE).
 SEQUENCE 2249 AA: 224333 MW: A9D6646C089DE087 CRC64:

Query Match 5.9%, Score 630.5; DB 1; Length 2249;
 Best Local Similarity 23.2%; Pred. No. 1.9e-16;
 Matches 500; Conservative 212; Mismatches 779; Indels 665; Gaps 109;

QY 180 LADTLMQYRTHAG-----HASTAV-----GMSYAG-----HFSNAFGTR- 218
 1 MANISPKLEKAIQGLAALFTSTAIMLSSGALGAVGATATNNAAFSNVNGNN 60
 QY 219 -----ATAEQT-----YSLAVLTATAKASSIVGSAQAGFAATVAGST 261
 61 WMEITRAGVANTPAGGQONMMAFYGGDYVTADADRIRIKAIN-----VAGTT 110
 QY 262 QVNLN-----RG-----IALGFSOVLQKNDVNAANVAVAPDDNPIDNRYK 305
 111 PVLGITONTVVGSITITGNLLPVLNAGKSLTLGNNAVANAGFADADNY----- 163
 QY 306 ATEKAGADVPSIGNSNDISIRKIIIVGAGSADTDVANAOLKEAVRLANROITPKGD 365
 164 -----TGLGIALGANAAL-----IIOGAAPSKITLAGNID-----GGGIIYATD 205
 QY 366 DSNNEVEKGLKRTITGGAQTSALTLDHNIYVONGDLKQOLAE-----TLTSLKMTT-- 420
 206 AAINITIGNTALATVNVGAGATATLG-----GAVIKATTTKLTNAAVLTITLNAVLTGA 261
 QY 421 -ENLTANEKVY---GKRLTLDKIGFTNDMNGIDESKPYLDKDTGTHAGGOKITKLA 475
 262 IDNTTGGNVGLNLNGALSOVTDIGNTSLATISVG-----AGATLGGAVIKATT 314
 QY 476 GVVDADAPYGLKKVNOT-----AESALQFTYKVKDNGKDADSDIITY--G 523
 315 TKLIDAAVKEFTNPVVYTGALDNTGNANNGIYVFTGNS--VTGNGVFTNLATVNVGAG 373
 QY 534 KNNKPDG--TOVNTLKLKENGVDVTTETNGVTEGL-----NONNGLT--VGNSTLNDG 575
 374 LLQVGGVAVKANTINL--TDNAAVTFPNPVYVGAIDNTGNANNGIYVFTGNSVITGD- 430
 QY 576 LSVKRTNSKQIQVG-----ADGIFTDISNSKPPAGJENTTRI 614
 431 --IGTNTLATVNVGAGATLGGAVIKATTTKLTNAAVLTITLNAVLTGA-IDNT-- 485
 QY 615 TFDGIFPANNNGSLDANKPRLPTPTINGAGKELTWNQSAIINPATNGGOLDPMNRSTANT 674
 486 GGDNVGVNLNALS-----QVTGNIGNTSLATISVAGATATLGGAV-----IKATTT 534
 QY 675 EKSQAATIKDLYNLQVPLTFAGDTGNVTKKLGELLKVGKGTTPADLTNN---IG 730
 535 KLTDAASVKK--FTNPVVYTGALDNTGNAN--NGIYVFTGNSVITGDIGNTSLATIS 588
 QY 731 VVADSTDSINPKLAKTSLDADAVTKLTITASDKV---TVDS--GNNTATLO--NGDITF 783
 589 VGAGATLGGAVIKATTTKLTNAAVLTITLNAVLTGAIDNTGQDVNVGLNLNGALS- 647
 QY 784 SKONTGATPATNS--KTIGVDGLKFTDNNGIALDGTYYTKKVGKAGKODGSLDKSKPYLD 842
 648 --QVTGIDIGNTSLATISV--GAGTATLGGAVIKATTTKITNAVASVK-----FT 693
 QY 843 KDKLVGEVEITNGIN-----AGKAITG--LSNLTIDAT--NATTVHVTQIGIV---DST 892
 694 NPVVYTGALDSTGNANNGIYVFTGNSVITGDIGNTALATVNVGAGATLGGAVIKATTT 753
 QY 893 DKTRASIGDVNAGFNKKNNGDAKDFSTYDVFINGNATTKATVD--GKASVADV 951
 754 KLTNAASVLTITLNAVLTG---AIDNTGNGVNLNNGALSQVTDIGNTSLA--TI 809
 QY 952 NVDTGTTILTGADGNKNOIGVKTTLTKTDKAGKAINFVSNGSGDKALINAKDIADNLN 1011
 810 SVGAGTATLGGAV--VIAKATTTKLT-----NAAVLTITLNAVNV----- 845

QY 1012 TLAGEIRNKGTADTLQOTQVKKVRENG--DDNDADTITVGEKDAK-----NOVNTLK 1064
 846 -LIGAVDNTTGGNVGLNLNGALSQVTDIGNTSLATISVAGATATLGGAVIKATTTK 904
 QY 1065 LKGRNGIDTQTKNDGYTFEGINTQSLKAGNNT--TLANNGL-----STKNTAGNEQIOY 1117
 905 LTNAAVLTITLNAVLTGAIDNTTG---GDNVGLNLNGALSQVTDIGNTSLATISV 961
 QY 1118 GA-----DGVKFAKVNNGVAG--IDGTRITRDEIGFAGTN---GS 1155
 962 GAGTATLGGAVIKATTTKLTDAAASVAKFTNPVVYTGALDNTGNAANGIYVFTGNSVITGN 1021
 QY 1156 LDKSKPRLSDKGINAGKKTITNIOSEIAONSDAVATGKIYDLKTELENKISSTAKTAQ 1215
 1022 VGNNTA--LATVNVGAG---LLQVGGVAVKANTINLND-----NASAVFT 1062
 QY 1216 RSLHEFVADQGNFTVSPSSYDTSKSDVITTAGENGIT-----TKVAKGV 1266
 1063 NPVVYTGALDNTGN-----ANNIGVTEGTGNSVGNVNTNALATVNVG-- 1106
 QY 1267 RVGIDQTKG-----LTPKLTVG-----NNGKGIYIDSONGQNTIT 1303
 1107 -AGLLQVGGVAVKANTINLTDNAAVTFPNPVYTGALDNTGNANNGIY--FTGNSVIT 1163
 QY 1304 G-LSNTLANVNDKGSYRTTEQGIIRDEKTRAASIVD-----VLSAGFNLQNGEAVD 1357
 1164 GDIGNTALATVNVGAGITLQAGSL-----AANNIDFAGSTLEFNGPLDGGKALP 1216
 QY 1358 F-----VSTYDTVPAGGNATTAIVTYDDTDSKTSKVY 1390
 1217 YERKGAIANGNMILNVNKTTLASHLTIGTVAEINIGAGNLETIDASVD-----VT 1269
 QY 1391 DVNVDDTTEIVKDKLGKKTITLTSTGANKFALSNOATGALVRAASIVD----- 1442
 1270 ILNAQNTINFAKRSVLTLSL---TGCVYNNILL-----AADDVAPGADGTV 1314
 QY 1443 ---HLNTLSGDIQTAKGASOANSAGYVDADGNK---VIYDS---TNKKYQAKNDGT 1491
 1315 VFNGGVGNL-----VGSNVAGTARNIGDGGNKFTLLIYNAVITTDVNLGEGIQNL 1368
 QY 1492 VDKTKVEAK-----DKLVAQAPDGLT--AQNMYKSVINKQEVNDA 1531
 1369 INKRADETSSTAFNAGAIQINDATYITIDANNGLNIPAGIOPAHADQVLQONSSGNR 1428
 QY 1532 NKKGINEDNAFYVKEKASDNKTKAAVTVGLDLNAV--AOTPLTFAGDTGT--AKRL 1587
 1429 TITIGANID-----PDNDDEGIYV--LMSVYAGKKTITG--GKTFGGAHKL 1471
 QY 1588 GEYLTITG--GOTDNTKLTDNNIGVAGTGFYKLAQDLNLNSVANAAGTKIDDKVSEV 1646
 1472 -QTLTFKAGDGSF-----AGTFNTNIVLIDT----- 1499
 QY 1647 DSSGOAKANPVLASNGLDGKGVISNVGKGTQDTPAANVQOLENVRNLLGLGNAG---N 1703
 1500 ---GQ-----LEIG-----ATTANVLENDVAVOLTQICNIGGFID 1531
 QY 1704 DNAD-----GNOVINADIKRDNSSGSSNRIVIKAGVILGKGNNDTEKLA---TGVO 1754
 1532 ENAKNGAVTLNNVNVVAGAV--NTGNTNGTL---VLGASNLNRVANGIAMLKVAGCN 1585
 QY 1755 YGVUDKONAN--GDLSNVWVATQDGSKKALLATYNAAGOTNITLNNPAIDRINQGR 1813
 1586 VTIKAGKVRIGEGGTGT---LTLPAHFNLTG-----SINTGQALK 1628
 QY 1814 FFHYNDNQPVPVQGRNGIDSSASGKHSVAIGQAKADGEAAVAIGROTQAGNOSTAIGD 1873
 1629 LNFPMNGSVSGV-----GTAANSVDITTAGTSPASSV 1663
 QY 1874 NQAT-----GDOSTA---IGTGNVAGKHSIGAIDPSTYKADNSYSVGNNOFTDATQTD 1926
 1664 NAKGTATLGGTSPANFTNTGAVTLAK-----GSITSFAKNTVAT 1704
 QY 1927 VFGVGNNTVYTESNSVALGSNSAISAGTHAGTQAKSDGTAGTTTGA-----IGTVAG 1981

DB 1705 SF-VANSATTFNSNSLAFNSN--ITGC-----GTLTGAGANVTTGT-GS 1746

QY 1982 FAGOTAVGAVSVC--SCAERRIONVAGEVSATSDAVNSOLYKATOGIANNTN 2035

DB 1747 FTDTLTUFTTDFGAAGAKSGCNLTIRSGSLDLSGVSTLAL-----VVTATN 1791

RESULT 2

OMPA-RICCN STANDARD: PRT: 2021 AA.

ID OMPA-RICCN

AC Q52657: P95591; P95592; P95593; P95594; Q52667; Q52668; Q52669;

AC Q52670: Q52674;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface antigen) (rOmpA) (rOmpA A).

GN OMPA OR RC1273.

OS Rickettsia conorii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiidae; Rickettsia.

OX NCBI_TaxID=781;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Malish 7;

RA MEDLINE=94171067; PubMed=8125327;

RA Croquet-Valdes P.A., Weiss K., Walker D.H.;

RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia conorii (Malish 7 strain).";

RL Gene 140:115-119(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Malish 7;

RA MEDLINE=21442074; PubMed=11557893;

RA Ogata H., Audic S., Remeis-Audiffren P., Fournier P.-E., Barbe V., Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M., Raoult D.;

RA "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.," Science 293:2093-2098(2001).

RL [3]

RP SEQUENCE OF 8-204 FROM N.A.

RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;

RA MEDLINE=97015921; PubMed=8862558;

RA Roux V., Fournier P.E., Raoult D.;

RT "Differentiation of spotted fever group rickettsiae by sequencing and analysis of restriction fragment length polymorphism of PCR-amplified DNA of the gene encoding the protein rOmpA.,"

RL J. Clin. Microbiol. 34:2058-2065(1996).

RN [4]

RP SEQUENCE OF 953-2012 FROM N.A.

RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;

RA Raoult D., Fournier P.E., Roux V.;

RT "Phylogenetic analysis of spotted fever group rickettsiae by study of the outer surface protein rOmpA.,"

RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.

CC -1- PTM: GLYCOSYLATED (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).

CC EMBL: 001028; AAA17405.1;

DR EMBL: AEO08674; AAL03811.1;

DR EMBL: UA3794; AAB49549.1;

DR EMBL: UA3798; AAB49550.1;

DR EMBL: UA3806; AAB49551.1;

DR EMBL: UA5244; AAB49566.1;

DR EMBL: UA6918; AAB86663.1;

DR EMBL: UB3440; AAC35176.1;

DR EMBL: UB3443; AAC35179.1;

DR EMBL: UB3448; AAC35184.1;

DR EMBL: UB3453; AAC35189.1;

DR InterPro: IPR003858; rOmpA_rOmpB.

DR Pfam: PF02708; rOmpA_rOmpB, 1.

KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein;

KW Complete proteome.

FT SIGNAL 1 38

FT CHAIN 39 2021

FT DOMAIN 238 946

FT DOMAIN 1424 1528

FT VARIANT 60 60

FT VARIANT 76 76

FT VARIANT 86 137

FT VARIANT 126 133

FT VARIANT 953 954

FT VARIANT 1245 1245

FT VARIANT 1308 1308

FT VARIANT 1877 1877

FT CONFLICT 10 10

FT CONFLICT 92 92

FT CONFLICT 126 126

FT CONFLICT 137 137

FT CONFLICT 157 157

FT CONFLICT 368 369

FT CONFLICT 374 388

FT CONFLICT 640 640

FT CONFLICT 669 669

FT CONFLICT 793 793

FT CONFLICT 803 804

FT CONFLICT 809 823

FT CONFLICT 898 898

FT CONFLICT 908 908

FT CONFLICT 985 985

FT CONFLICT 1009 1009

FT CONFLICT 1013 1013

FT CONFLICT 1182 1182

FT CONFLICT 1314 1314

FT CONFLICT 1451 1451

FT CONFLICT 1624 1624

FT CONFLICT 1628 1628

FT CONFLICT 1872 1872

FT CONFLICT 1875 1875

FT CONFLICT 1878 1879

FT CONFLICT 1936 1936

FT CONFLICT 1936 1970

FT CONFLICT 1997 1997

SEQUENCE 2021 AA; 203328 MW; 327FCA2D7CB24668 CRC64;

QY 14 GEFMAVAEAKSHSTGGSCATGQVSVFTLSFARIALAVLVGA-----T 60

DB 88 GDTTADADADHITAINVADTTPG-----LINAONTVGSIVGNNLPTITAGKSLT 143

QY 61 LNSAVVAGIGISEADGKGGANARGDKSIAGDIA-----QALSSQSTAIQDNKIVHNSN 115

DB 144 LNS-----NNADAANHGFAPADNTGCLNIALGAMNALLIIOSPAAPAKITLAGIN 195

QY 116 NNANIGAKASGHSIAIGG-DVLASGHASIAIGSDDLTKREYQOISELLPIRGOKAL 174

DB 176 GGGITVKTDAINGTIGNTALATVNVGAGIATVLEGAITKATTTKLTMAASVL----- 249

Query Match 5.1%; Score 550.5; DB 1; Length 2021;

Best local similarity 21.5%; Pred. No. 1.6e-13;

Matches 511; Conservative 243; Mismatches 811; Indels 811; Gaps 116;

QY 175 NDYQADPNLQKRRTHAQHASTAVGAMSAKGFHFNAGFTR---ATAEGTSLAVG 230
Db 250 -----TLTVNAVLGADINTGTVDNVGLNLNGALSOYTG 285
QY 231 LPTAKAASIAVGSNAQAIGFAATAVGSGTOVNLNRGIALGFGSOYLQK---INDVNA 287
Db 286 NIGNTNALATISVAGKATLGGA---VIKATTTKLDNASAVTFNPPVVTGADINTGNA- 342
QY 288 NVRATAPDNDPIDNRKATFNKATDVFSIGNSGNDISIRKLIINGAGSAD----- 340
Db 343 -----NNGIVTFGTSTVGNIGNTNA-----LATISVAGKATLGALIK 383
QY 341 -----TDVAVNAOLKEAVRLA-----NROIFGDDSNR-----RVEKG 374
Db 384 ATTKLTLDNASAVTFNPPVVTGADINTGNANNGIVTFGTSTVGNIGNTALATISVG 443
QY 375 LKLTLLTGG---AQSALTJHNGVONGDLKVOLETLISLKMVT---TENITAN 426
Db 444 AGKA-TLGGALIKATTTKLTLDNASAV-----TFNPPVVTGADINTGNAN-N 488
QY 427 EKVYVGTGRLLTDKIGFNDMGIDESKPYLDKDTGHAQ---GQTKTLTACVYDDA 482
Db 489 GIVTFGTSTVGNIGNTALATI-----SVAGKATLGALIKATTTKLDNA 537
QY 483 ATYGOLKRYNOT-----AESALQFET-VKVDKNGNDANDSKLITVGNKRP-DGTQ 532
Db 538 SAVTFNPPVVTGADINTGNANNGIVTFGTSTVGNIGNTALATISVAGKATLGAI 597
QY 533 VNTLKLKENGVDYTFENGIVTPEL-----NONNGLT--VGNSTLNDGSLVKNKTSNK 585
Db 538 IKATTTKLDNASAVTFNPPVVTGADINTGNANNGIVTFGTSTVGNIGNTALATI 654
QY 586 QIOVGA-----DG-----ITFTDISNSKPGAGIENTRITRDGIFANN 624
Db 655 TYNVAGAGIATLEGAVIKATTTKLTLDNASAVLTLTNNAVLTGA-IDNTTGV--DNVGVNLT 711
QY 625 TGSID-----ANKPRLPTGJNAGSKELTNVQSAI--NPATNGQL 663
Db 712 NGALSQVNTGNIGNTALATISVAGKATLGAVIKATTTKLTLDNASAVTFNPPVVTGAI 771
QY 664 DEMANLSTANTNEKSSAATINDLYLSOVLPLFAGDGPNTTKLGE-----I 711
Db 772 D-----NNGNA--NNGIA-----TFGTGS--TVTGNIGNTALATVNVGAGL 809
QY 712 LKVGSGKTADDLTKNNIGNVADSTDNLSJLVKAKTISDLDAVNTKLTASDKYTVDSG- 770
Db 810 LRVQG-----GVVKSNTIN-----LTD-----NASAVTFNPPVVTGAI 843
QY 771 NNTAKLQNGDLTFESKQK---TGATPATNS-KTIVGDLKFTDNGIALDGTITYTKDKYGF 827
Db 844 DNTGNANNGIYTFGTSTVGNIGNTALATISVAGKATLGALIKATTTKLTLDNA--- 900
QY 828 AKQDSILDKSPYLDKDLKAVGEVEITNGINAGKATISNLTLDATNATTHAVTQGL 887
Db 901 -----SAVTFNPPV-----VVTGADINTGNANNGIV-----TFGTSTVYG 936
QY 888 IYDSTDKTRASIGDVLNAGFNKNGN---DAK--DEVSTYDVFINGNATTAATYATYDGK 943
Db 937 NIGNTNALATVNVG---AGVTLQAGSLDANINIDF-CARSTLEP-NG-----PLDGG 983
QY 944 ASKAVAYDVNVGDTIHLTG--ADGNKNOIGVKTTLTKTDAKGDALNFVSNSGD----- 996
Db 984 GNAIY-----YFGAIALNGNALLNVNTKLTLYVHLITITVAEINIGAGNLEFAID 1034
QY 997 -----DKALINKADIADNNTLAGEIRNTKGTADALQFOVKYKRENDDDN---DADTI 1049
Db 1035 ASAGVYTLNMADI--HFRAL-----DSALVLSNLTGCVN---NILLADLV 1077
QY 1050 TVGKDAKTNOVNTLKLKKNGLDIQNTKGTVTFGINTQSGLKAGN--NTTLNNGLSIK 1107
Db 1078 APGVDEGTVEFD---GGVNGLNGSNAVAGAA-----RNIGDVGKKNKFNLLITNAVYIT 1128

QY 1108 NTAGNEQIOYAGDVKFAKVNNGVAGIDGTRITRDEIGFAGTNGSLDKSPHLKSDG 1167
Db 1129 D-----DYNLEIGIQVNLIN---NADFTSSTAFNAGTQIQUIDATYTTIDANNGNIN--- 1175
QY 1168 IMAGKRTINIGSGEIAONS--ND-AVTGKIYDLTELENKISSTAKTAONSLEFSA 1224
Db 1176 IPAGNIFAHADQOLILQNSGNDRTTLIGANIDPPNDDEGIVILNSVTA----- 1225
QY 1225 DEQGNNTVFNPPSSYPTSTSVITTFAGENGITTVKNGCVAVNGIDQTKGLTPKLTVG 1284
Db 1226 ---GKLTJIGKTFGGAHRLQDIY-FKGE-----GDFGAGTFE----- 1261
QY 1285 NNGKGIIVDSONGQNTITG---LSMTLANVNTDKGSVTFTEQGLIKEDXTRAASIYD 1341
Db 1262 ---NTTNIVLD-----ITQLELGATTAIVVLEKAVOLTOQG----- 1296
QY 1342 VLSAGFNLQNGEAVDFVSTYDTPVNFADGNATTAATYTDTSKSVYDVVNDVTTEY 1401
Db 1297 -----NIGG-----FLDFNAKNGVTTLNN-----NVNAGT--- 1322
QY 1402 KDKKLGVKTTTLTSTGT-----GANKFALSQATGDALVFRASDIVAHLNLTLSGDIOTAKGA 1457
Db 1323 -----VKNTGTINNGTLVIGASNL---NRVNGIMLKYG-----AGVNTIAGK- 1363
QY 1458 SOANSSAGYVADQNKVITYDSTDNKRYQAOKNDGTVDKTEVAKDKLVQAQPTPDGLAQM 1517
Db 1364 -GNVKIGEIOGTGNTL--TLPAHF--KLTSINKT-----GGQALKL 1401
QY 1518 NVKSVINKEQVNDANKKOGINEDNAFVKGLEKASDNKTKRNAVYTGDLNVAQPTLTRA 1577
Db 1402 N-----FMNG-----GSVGVVGTANSV 1420
QY 1578 GDITGT--AKRIGETLITKG---GQTDJTNKLDNNIGNVAGTGTFTKLAKDLT----- 1626
Db 1421 GDITTAATSPASSVANAKGATIGGTTSPAHFTFNNGAVTLAKGSTTSRAKNTATSAFA 1480
QY 1627 NUNSVNAG---GTYKIDKGV-----SEVDS-----SQOKANTPVL 1659
Db 1481 NSATINFGNSIAFNSNTTSGTLLTGLANQVYTGTSFPTDLTLTTPDGAASGGINLT 1540
QY 1660 SANG--LDGKGVISNVGKRTKTDANVOQLNEVNLGLGAGNDMDGQVNTADIK 1717
Db 1541 IKSSTLIDLSG--VSNIALVYTA-----NFDNM----- 1567
QY 1718 KDPNSSSSNRVYIKAGTYLVGKG-----NND-----TEKLIATG 1751
Db 1568 ---NISPDTKYVISAETAGLKPTEKENVKIRINDNRPVDFTPDASTLTLFAEDIAAG 1624
QY 1752 GVQYGVNDKGNANGDLSNV-----WYTKQDGSK-----ALLATYNAAGOT 1793
Db 1625 ---VIDEDFAPGGLANIPFANINIKKSLMEDADENGSDARQAFNFGMLPLOGADAT 1680
QY 1794 NYLNN---PAAEIDRINEGI-----PEFHVNDQCEPVYOGRGISAS 1837
Db 1681 THLMQDVYKSDITIAAVNNQVYASNISNTTALNAMAQKQVQAGNKPVSSEGDMD- 1736
QY 1838 GKHSVALIGFAKADGEAAVAIGRQTAGNOSTAIGNAOTG-----DOSIAIGTG 1888
Db 1737 -----AKFGAMISPVEGNATOKMCSIS-GYKSDTTGGTIGFDGFSVDLVLGLA 1785
QY 1889 NVYAGK---HSGAIGDPSYKAD-----NSYSVGNNNOFTLA----- 1922
Db 1786 YTRADYDILKNNKKTGDKNKNVESNIISLYGLVSPYENLEVEIASYSPDKTRSSRYI 1845
QY 1923 ---TQDFGVGNNTIYTESNSVAL-----GSNSAISAGTHAGQAOKSDGTAGT 1969
Db 1846 ATLETFVGYTFANGKYSSEYTGQOLMAGYTYMSEMINLPLAGLRYSTIKKSYKEJCT 1905
QY 1970 TTTAGATGVKFGAGOTAVGAVSVGASGAERRIQNVAAGEVASTSTDVANGSOLIKATO- 2028
Db 1906 TY---QNLTVKKNYTFDGLGAKVS-----SNINVEIYVLT-----PELYAMVDY 1949
QY 2029 GIANATNEIDHRIHQENKRNANAGISSAMAMASMPQA 2064


```

      | : | | : | : : | : |
Db 1950 AFKNKVSARDLQ-----GMTAPLPNTSEKQS 1977

```

RESULT 3

ID YDBA_ECOLI101 STANDARD: PRT: 2003 AA.
RX P33666: P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ydbA.
GN ydbA OR B1401/B1405.
OS Escherichia coli.
OC Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae:
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Petráň N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RL "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Maki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubramanian S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horichi T.;
RL "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map."
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 464-2003 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92190338; PubMed=1665988;
RA Moszer I., Glaser P., Danchin A.;
RL "Multiple IS insertion sequences near the replication terminus in
RT Escherichia coli K-12.";
RL Biochimie 73:1361-1374(1991).
CC -! SIMILARITY: TO S.TYPHIMURUM ORF NEAR CYSG (AC P25928).
CC -! CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
CC THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
CC BETWEEN AMINO ACIDS 839 AND 840.

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).
CC or send an email to license@isb-sib.ch.

DR EMBL: AE000237; AAC74483.1; ALT.SEO.
DR EMBL: AE000237; AAC74487.1; ALT.SEO.
DR EMBL: D90778; BA15009.1; ALT.SEO.
DR EMBL: D90778; BA18880.1; ALT.SEO.
DR EMBL: D90779; BA18881.1; ALT.SEO.
DR EMBL: X62680; ? NOT_ANNOTATED_CDS.
DR EcoGene: EG11307; ydbA.
KW Hypothetical protein: Complete proteome.
FT CONFLICT 489 489 I -> V (IN REF. 2).
FT CONFLICT 495 495 I -> V (IN REF. 2).
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;

```

Query Match      : 5.18; Score 544; DB 1; Length 2003;
Best Local Similarity 21.38; Pred. No. 2.7e-13;
Matches 415; Conservative 242; Mismatches 695; Indels 600; Gaps 102

```

| | | | |
|----|------|--|------|
| QY | 274 | GSYOLODNQVYNAANVAAPADNOPIDNRKYA-----TFKKCATDVFISGNSNG | 325 |
| Db | 122 | GDEITIPDDPDTPEPKPEVSEFNDVILDTKEKLTITRDSVFYTTENADSTISLODSNGK | 181 |
| QY | 326 | SIRKRIINVGSGSADTDAVVAQLEKAVRLANROITFKDDSNRVEKAGKTITITGSA | 385 |
| Db | 182 | A-----TIMQIDE-----ANNVALEG-----VSADG | 205 |
| QY | 386 | QTSALTHNIGVONGGGLVQLAELTISLKMATTENLTANKEYVGRLLTTD-KIGFT | 444 |
| Db | 206 | ATKMOYHNHELVTYTG-----NATVNNN---GKT---TVDKSDSTG | 241 |
| QY | 445 | NDMGIDESKPYLDKPDGJHAGGOKITKLTAGVDDAATYQOLKVKNOJASALQTFV | 504 |
| Db | 242 | TEING-NGKVIQDDGDLVSGGCHID-----ITGDSATVDN-----KGTMY | 283 |
| QY | 505 | KKVDKNGNDANDSKITIVGRKNRPDGTQVNTLKGENGVDYTTETNGTVTEGLNONNGL | 564 |
| Db | 284 | JDPESMGIOIDGDKAIY---NNEGSEITINGGTGTOINGDDATANNNGKTVDGKSDTG | 340 |
| QY | 565 | TVGSTLNDGLSKKNTNSKKOIQVGADGTTTFPDISNSKPGAGIENTTITITDGI GFAN | 624 |
| Db | 341 | EI-----NKNKVIQDG-----DLVSGGCHID-----ITGDS-ATVDN | 375 |
| QY | 625 | TGSLDANKPRLTPGIAGCKELTNVOSAINPATNGGOLDPMRLSTANTKESGSAATIK | 684 |
| Db | 376 | KGTMVTVDPPE-SIGIOVDDDOAVNNNEGSATITNG-----TGTQINGDADATAN | 423 |
| QY | 685 | DLVMSOVPPLTFADDTGPNATKKGELITKYGKGTATADLTKNKINGVADSDNSLTVKL | 744 |
| Db | 424 | ---NNGKTVYDGKSDSTETIAGNNKVIQ-----DGLDVLVSGGCHIDITGDSATV- | 471 |
| QY | 745 | AKTISLDADVNTKRLTASDKVYVDSGNTPAKLON--GDLTFESKONGA-----TATNSK | 797 |
| Db | 472 | -----DNKGTMYVDPESIGIQIDDDQIVNNESESTINGGTQTOINGDADATANNNG | 524 |
| QY | 798 | TIGVGLKFTPDNNGIADGTTYITKDKVGFAPKODGSLDKSPYLDKDKLVEVEITING | 857 |
| Db | 525 | KTYVDG---KDSITCKTAG-----NIGVNLNDSL-----TYVCG | 556 |
| QY | 858 | INAGGAITGLSNTLTDATNATTGHTVTOIGIYDSTDKTRASIGDVLNA-GENLKNNGDA | 916 |
| Db | 557 | AH-----GVENIGDNGTNNKGD---VVSDT-----GSIGVLINEGATVTSBGV | 600 |
| QY | 917 | KDEYSTDYDVFINGNAFTAKVYTDGKASKVAADVAWDGTTIHLTGAD--GKNKOIGVKT | 974 |
| Db | 601 | N-----VSNEATGFSITTNSGKVSLASMOVGPDS---TGVDLINGNNNSV---- | 642 |
| QY | 975 | TTLTKTDAK--GDKAIMEVSNGDDKALINAKOIAONLMTLAGEIRNTKGTADTALQTFQ | 1033 |
| Db | 643 | TLAKDLKLVGCKATGAINV-SGD-----ANTVITNGVVLVDKXKTRADNAEYEF | 690 |
| QY | 1033 | VKKAYKEN--GDDNDADTTVGDAKACTNOVNTLKLKGKNGLOIQNK-DGYTVPFGINTOS | 1088 |
| Db | 691 | DPSGVINWYGSDDN-----VTLDGKLTIVYDSEKVSIR---QSMLEFGSA-----EKTS | 735 |
| QY | 1090 | GLKA-GNNTLINNG-----LSIKNTAGNEGIOVGAGGVFAKVNNG-----VVGAG-- | 1133 |
| Db | 736 | GLVYIGGNTVYNNANGGELIGEKKAL-----AGSOVTSLSRTGYSTTVSYIVVSGES | 787 |
| QY | 1136 | ---IDGTRITPD-EIGFAGTNGSLDKSPHLSKDGINAGGKKTITNIOSEIAONSDAVT | 1192 |
| Db | 788 | VYLNQDPTTISEPFLGPAV-----IRYODKALLEITGSG----- | 821 |
| QY | 1193 | GKGIYDTELEKNKISSATATAONSLEHFSVADDEGNNFTVSNPYSYDTSKTSYD--IT | 1250 |
| Db | 822 | -----ATLTMDODDISFHHG-----TTFVELOING | 846 |

Db 247 NATANALNLOAGGT-----INFNGIDGTGRVLILSKNCAATDFNVTSIGSGLNK-GI 299
 QY 851 VEITTINGINAGKAITGLISLTLDATNATGHTVQLGIVSDTKTAAST-GDVLNAGFN 909
 Db 300 IELMTVALINGOLINAGPANAIVGTNN--GAGRAAGFVSVGDKAKATIDGQY----- 351
 QY 910 LKNKNDADPEVSTIDYDFINGNATKATYTDGKASKAYAVYVNDGTTILHTGADGKNQ 969
 Db 352 -----AKDNY-----IOSANAN--GOVNF-----RHIVDVGIDGTTAFKTA-- 386
 QY 970 IGVKTTTLTKDAGDKALINF-SVNSGDKKALINAKDLADLNLTLAGEIRTKTADTAL 1028
 Db 387 ---SIVAITONS-----NEGTTDFGLAAOVYPPDMITLTGNFTGDANNPGTA--GV 434
 QY 1029 QTFQYKRVKRENGDDNDADITTVGDKAKTNOVNLKLGKNGLDIDTKKDGTVTEGINTQ 1088
 Db 435 ITFPAANGTLASASADANY-----AVTNMTITAIASGVGVVQLSGTHAELRLG--NAG 485
 QY 1089 SGLRAGNNTTLNNGSLIKNTAGNEQIOVGADGVKFAKVNNGVYAG--IDGTRITRD 1145
 Db 486 SVEKTLADGTIVNGK-----VNGTFLVG-----GVLAAGAITLIDGSAITITGD 526
 QY 1146 EIGFAGTNGSLDKSKPHLSKDG--INAGKKITNIQSEIAONSNDVATGKTIYDKTE 1202
 Db 527 -IGNGGGGALLOST--LANDATKTLTGANITISANGTI--NFGANGTI----- 573
 QY 1203 LENKISSTAKTAQNSL--HEFSVADEQNNETVSNPYSSYDTSKTSQVITFPAENGIT 1259
 Db 574 ---KLTST---QNNIYVDCDLAIATDQTVGVDASS-LTNQTLTISGLIGANNTTL 625
 QY 1260 KYNKGVYRVGIDQK-----GLTTPKLTIVGNNN-----GKG-IVIDSQ 1296
 Db 626 ---GQFNIGSSKTLNGNVAINELVINGNSVOFAHNTYLLITRTNAAGKLTIEPV 681
 QY 1297 NQONTITGLSNTLANVTDKGSVTRTEOGKTIKDEKTRASIVYLSAGFNLONGEAV 1356
 Db 682 VNNNTTLAAGTILGSAANPLAINGSKG-----ARADVLAV-----GECV 723
 QY 1357 DEPVSTYDV-----NFAQGNATYAKYVYDDTSKTSKYVYDVNDTTIEVKDKL 1406
 Db 724 NLYATNITTTDANVSFEVFNAGKNIIVSGTVGGQGNKFNVALD---NGTIV---KFL 776
 QY 1407 GYKTTTLTSTGANKFALSNQATGDALYKA--SDIVAHNLNLSGDIQAKASQANS 1464
 Db 777 GNATENGNTTIANSTLOISGNYTADFLASADGTGIVEFNTPINVLTKAKAVPNAL 836
 QY 1465 GY-VIADGKVIYDSTDNKYQAKN-DGTVDKTEKVAQKDLVAQAQTPD-----GTIA 1515
 Db 837 QITVSPGNVYV-----NEIGNAGNYHGAMPTDIAFENSLSGAVLELPBGIFPNAGNTI 891
 QY 1516 QMNVKSVINKEOVNDANKKQGINEDNAFYKLEKASDNKT--KNAAYTVG--DLNAV 1569
 Db 892 PLTIKSTVGENETA-----EGFSVSPVIAVSGVSYADQVIGDQNNITGLISGNGII 945
 QY 1570 AQTPLTFADDTGTAKKIGETLITLIGCGDPTKLDNNINIGVAGVJDCG--TVLARDLT 1628
 Db 946 VNATTLTYAG-IGTINNQG--TVTLISGVPNTPG--TVYGGTIGISKEQVFTTIDYNNL 1002
 QY 1629 NSVNAAGCTIDD-----KGVSFVD-----SSGOAKANTPYLSANGL----- 1664
 Db 1003 GNIAIANTTINDGVYTTGGIAAGTAGIDPQKILTLGSVNGANANRFPADGIFSNSTMI 1062
 QY 1665 -----DIGKVISNKGKGTQDPAANYQOLNEVRNLGL--GNAGNDNADGNQYV 1712
 Db 1063 VTTKANNGTIVTYLGNAFVGNI--GSDDTFVASV--RFTGSNNGAGLKGNTYSOVIDFGTVN 1119
 QY 1713 IADIKKDPNNGSSSNRFTVAKG--TVLGGK-----GNN-----DREKLA 1749
 Db 1120 LGIV-----NSNVILIGSTTAINGKIDILLNTLTFAGSTWGNNTSLETTTLTA 1169
 QY 1750 TG-----GVQGVVDKDG-----NANGDLSNVVYKTKQDKSKR----- 1781
 Db 1170 NGNIGHIVIAEGQVAVATTTGTTINVDNANANFSGTGYTLTIGGAPRNGTIGPNFT 1229

QY 1782 -----ALLATYNAAG-OTNYLNNPAAEIDRLNKGIRFEH-VND 1819
 Db 1230 VTGNSREPVNYGLIRANQDYVITRTNNAENIYVNDITNSPFGAPGVGQNVTFVATMT 1289
 QY 1820 GNOEPVVOGNGIDSSASGKHSVAIGFOAKADEAAVAIGROTQAGNSIAGDNNQATG 1879
 Db 1290 AAYNNILLAKNSDASNFEVGTIYTDISAATTNQDLVAKDIOQLGNRLGALK----- 1342
 QY 1880 DQSTAGTGNVAVAKHSGAIGDPSYKA-----DN-SYSGVNNQTTDQTDVPGVGN 1933
 Db 1343 ---YLCTPEMV--GSEAGAI--PAAVAGDEAVDNNVAGTIAKPFYDHAQSKKGL--- 1392
 QY 1934 IYTESNSVALGNSALSASQTHAGTOAKKSDGAGTTTATGATGYKFGAGOTAVGAVSY 1993
 Db 1393 -----AGYKAKTGYIYGLDTLAN-----NNLMIGA-AI 1420
 QY 1994 GASGAERIRIONVAAGEVASTSDAVNG--SOLYKATQIGIAN--ATMELDRHIOENK-- 2047
 Db 1421 GITTDTLKHQDYKGD-----KTD-VNGFSPLVGAQGFENFPAGSAIFSLNQVKNK 1475
 QY 2048 -----ANAGISSAAMAMAPQATIPGRSMYTGIAI-----HNGGAVAV-----GLSKLSD 2094
 Db 1476 RYFEDDANGNMSKOIAAGNYD-----NMTEFGNLTGYDYDNAMQGVLYTPMAGLSYLS 1528
 QY 2095 NGQWVEKING-----SADTQGHVGAAYGAG 2119
 Db 1529 SDE-NYKETGTVANKOVNSKFSRDTLIVGAKYAGG 1564

RESULT 5

BICA_SALTY STANDARD; PRT; 1953 AA.

ID BICA_SALTY STANDARD; PRT; 1953 AA.
 AC P25927; P25928; Q9XQ03;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE putative surface-exposed virulence protein bigA precursor.
 GN BICA OR STM3478.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SCSG1412 / ATCC 700720;
 RA MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sandersen K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Potwolk S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2.";
 RL Nature 413:852-856(2001).
 RN [3]
 RP SEQUENCE OF 1-765 FROM N.A.
 RC STRAIN=LT2;
 RA MEDLINE=91100301; PubMed=1987123;
 RA Wu J.Y., Siegel L.M., Kredich N.M.;
 RT "High-level expression of Escherichia coli NADPH-sulfite reductase:
 RT requirement for a cloned cysC plasmid to overcome limiting sthC
 cofactor.";
 RL J. Bacteriol. 173:325-333(1991).
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
 CC in positions 414 and 732.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

[illegible]

Query Match 4.68; Score 493.5; DB 1; Length 1953;
Best Local Similarity 20.68; Pred. No. 2e-11;
Matches 431; Conservative 244; Mismatches 742; Indels 677; Gaps 100;

[illegible]

| | | | |
|----|------|--|------|
| Db | 571 | DNLSLQADDLVLTGAGMIITYGNGEAKNTG--NATVRADSVGF----- | 615 |
| Qy | 635 | LUPITGIMNGGKRELIVQASAINPATNGOGLDMNKLSTANTEKSSAATIDLYNLNQVPL | 694 |
| Db | 616 | -----VVAEKK-----NFKKKGGID-----VSLNLTGALVSG--DMSQVYL | 650 |
| Qy | 695 | -----FPAADGTPNVTKKLGEILKKGKKTADDTLTKNNIGVANDSTNSL | 740 |
| Db | 651 | DDGIMVSVQDESEGVSSATSVAS-----GDSNAVDIT--GNVIMASIDYQDDL | 698 |
| Qy | 741 | TYKLAKTLSDDAVNTKRLTSLASDKYTVDSGNNTKKLONGLTFSKQNTGATPATNSKTTG | 800 |
| Db | 699 | -----AACAPPLTG--VVGNGNGNTVTL--NGALNIDDDNLS---ATGQOYLD | 739 |
| Qy | 801 | VQGLKFT--DNNGIALDGTFTYITKRVGFARQODSGSKPYLMDKRLKJGEVETITNGIN | 859 |
| Db | 740 | VVGSLTSGDDNDVEIDGGINTH-----SEDLPLDGSADIT--GIS | 778 |
| Qy | 860 | AGGKATIGLISNLTDAINATLT--GHVTOLGIVDSTDKTRAASICD--VLNAGFLNKNNGDA | 916 |
| Db | 779 | VSGNSTVTJLNGHSITDFTNVVGHV---VLARNNNGSILTDGDSVVDVNVSYIPTG-- | 832 |
| Qy | 917 | KDFVSTYTFVDFINGNATTKA---VTYDQKASVAVD-----VNDGTTIHLTGADGKK | 967 |
| Db | 833 | -----YTYAALMADEGEGSLENKCDITISHGVYSYIRADNGSEVNSGDIILVATSNSSSE | 889 |
| Qy | 966 | NOIGVKTTLTKTKADGKKAINFVSNGDDKALIN----- | 1007 |
| Db | 890 | DR-----AATIRASEGSVAHN--KAGDITILISDTPGSGGIEVYPLKMYHTHFYAM | 941 |
| Qy | 1003 | -AKDIADNLNLTLAGIIRTKCTADTALQTOYQVKVRENG-----DDNDADPTIV | 1055 |
| Db | 942 | MASDGDVAVND--EGATJTLHOGAGYGYATSRGALNNGNTYIDGLPTLDEENNTTSY | 1000 |
| Qy | 1052 | GKDAKTNQVNTLKLKGKNGCIDQTKNKQGTFTFGINTQSGLKAGNNTTLNNGSHIKNTAG | 1111 |
| Db | 1001 | W-----QPSLYLTSSGVMVAGSIDADGAF--AINT-----GNTYVNNAGGMALNG | 1046 |
| Qy | 1112 | NEQIOYGA-----DGK-----FAKNNNGVCGAGIDGTRITRDEIGRATNGSLD | 1157 |
| Db | 1047 | GTAINGQVITLTADGVTGAGADELVGMAALNGVAVINDSGVINIDAD---YGOAFILSD | 1102 |
| Qy | 1158 | KSKPHLSKDGINAGKKITNQSGLIOMNSNDAPVATGKIYDKTLELNKISSRTAQAONS | 1212 |
| Db | 1103 | SSSYIINNNGSINLNGSM-----DOTDSHMG-----FTPDKATWOS | 1133 |
| Qy | 1218 | LHESVADQGNNTFVSNPYSSDTSKTSVITTF--AGENGITTKVKNKGAVRVG--IDQTK | 1274 |
| Db | 1140 | L-----PESGSDSDFRTPSGTFPAG---TLANYGTEJLNGDVWNG | 1177 |
| Qy | 1275 | GLTTPKLTGVNNGKGIIVIDSOMQNTITLSMTLNVNTDKGSVATTQCKIINDEKTI | 1334 |
| Db | 1178 | G-----WLYNEGASLTV---NGYITINGCAALAN----- | 1205 |
| Qy | 1335 | RAASIVDVLISAGFNLONGEAVDFVSITYDV--NFADGNATLT-----AKVY-----DDTS | 1383 |
| Db | 1206 | -----YETLDA--DAISTHSLFNEADGSIITPDLTLNGLVGYEYNNNGDFTG | 1249 |
| Qy | 1384 | KTSKVVD--VANDDPTIEVKKKL-----GKTTTLTSLTGANKFKLSQATGD | 1422 |
| Db | 1250 | SIATGTSQOEIVNTGDMTVAEDKSLVSGSEFFYNEEDATLITNLSGAVE---GGEENTII | 1305 |
| Qy | 1433 | ALVAVSDIVAHLLNTLSGDIOTAFKAGASOANSAGVYVADGNKVIYSTD----- | 1480 |
| Db | 1306 | NLVRANDSLQOVN--SOTIATNGSBAITTYVNSNDP---KMIWMTATGVINGINIPDAPL | 1360 |
| Qy | 1481 | ---NKTYQAKNDGTVVDKTKVYAKDKLVQAQITPDGTLAOMNKSVINKEQVN---DANK | 1544 |
| Db | 1361 | INLGRGNFNGOGITIN---VOGDNAVA---ISGTSYSY--VINLVNGLTIVGTEQKE | 1411 |
| Qy | 1535 | QGINEDNAF--VKGEKVASDNKTKRNAAVYVGDNLVNAQPIPLFEAG----- | 1578 |
| Db | 1412 | DGTNGTGLITGKNGNATITNNIAD---GYINVYADDSIAFGKTYAIIINGEINILLC | 1466 |

QY 1579 DTGTTAKKLGELITKGGOTDNKLTNDNT--GVYAGTGTGTVKIAKD-----LTN-LN 1629
 DB 1467 DSGCIVAPGTT-----GTQNHNGTADIVIPDATTAPTEGSIPTPPADPNAPQOLSNYIV 1522
 QY 1630 SVNAGGTIDKXGVSVDSGQAKNTVPLVSLNGDLGKYSVNGKSTKPT-----1681
 DB 1523 GTNAG-----SSGTLKANNLVIGDN-----VKYDTGFTSGTADITYVVDNAF 1565
 QY 1682 DAANYOOLNEVRNLGLGNA-GNDNADGN-----OVNIADIKKDPNSGSS 1725
 DB 1566 TGSNIQADNITSTSVVNNAGSQDADGVDTMTKNAYADVATDSVSDVAQALDAGYT 1625
 QY 1726 SNR---TVIKAGIVLGKGNDEKATAGVGVYVCKDNAGNDLSNNVWKYQKDSKAL 1783
 DB 1626 NNELYTSILWGT-----TAEINLSALKOV-----SGAQTATVFEARVILSNRFTM 1669
 QY 1784 LATVYANAGOTNYLTNNPAEADIRINEGIRFPHVNDGNOEPVQGRNGIDSSASGKHSVA 1843
 DB 1670 LA--DAAPQ-----IKDG-----1680
 QY 1844 IGFQAKADEEAAVAIGROTQ-----AGNQSIAI-----GDNAQATGDSI 1883
 DB 1681 LAFENVYAKDPPRAELGNPTQYDMLAROTLDLTASQNLTEVGIARLDGDSKTAGDNGL 1740
 QY 1884 AIGTGVNVAKHSAGIADPSTYKADNSVSVGNNOFTDATQTDVGVGNNTVTSNSVA 1943
 DB 1741 TCGYSOFELGKHSMAFDEGLAWNLSLRVDVHN-----LDSSRSVAVADVKIADSDMRQY 1796
 QY 1944 LGSNSAISAGTHAGTQAKKSDGTAGTTAGTGVKAGAGTAVGAVSAGSAGERRIQ 2003
 DB 1797 L-----ERSEB-AKFTTMMGDALVTPYAGKFKHTMEDGKESAGDF 1840
 QY 2004 NYAGEVSATSTDVANGSQL-YKATQG-IANATNEIDRHONENKANAGISSAMAMSM 2061
 DB 1841 NLSMNSGNFTAVDSITVGLKLDYAGKDGWATPTEGGPRLYSKQORFASLGGAAG-----1896
 QY 2062 PQATYIPGRSMVTGATTHNGQAVAGVLSKLSDC-----QWFKNGSAD 2107
 DB 1897 -OSFVGDGOKGGGV---NGLATIGV---KYSSNDTALHLTDAYQW-KEDGISD 1941

RESULT 6
 OMPB_RICPR STANDARD: PRT: 1643 AA.
 ID OMPB_RICPR 053020: 092CM0:
 AC 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (OMPb)
 DE (OMPb) (Contains: 120 kDa surface-exposed protein (Surface protein
 GN OMPB OR SPAP OR SPA OR RP704.
 OS Rickettsia prowazekii.
 OC Bacteria: Proteobacteria: alpha subdivision: Rickettsiales;
 OC Rickettsiaceae: Rickettsia:
 OC NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-BREINL:
 RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
 RA "Characterization of the gene encoding the protective paracrystalline-
 RT surface-layer protein of Rickettsia prowazekii: presence of a
 RT truncated identical homolog in Rickettsia typhi.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BREINL:
 RA Moron C.G., Yu X.J., Walker D.H.;
 RA "Sequence analysis of ompb of Rickettsia prowazekii.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBD databases.

RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MADRID E.
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Minkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RL Nature 396:133-140(1998).
 RN [4]
 RP PARTIAL SEQUENCE.
 RC STRAIN-BREINL:
 RX MEDLINE=92114896; PubMed=1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RA "Mapping of monoclonal antibody binding sites on CNBR fragments of
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
 RT prowazekii.";
 RL Mol. Immunol. 29:95-105(1992).
 RN [5]
 RP IDENTIFICATION OF CLEAVAGE SITE.
 RX MEDLINE=92104668; PubMed=1729180;
 RA Hackstadt T., Messer R., Cieplik W. Jr., Peacock M.G.;
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
 RT membrane protein of rickettsiae: identification of an avirulent mutant
 RT deficient in processing.";
 RL Integ. Immun. 60:159-165(1992).
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 CC LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPb FAMILY.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC CC
 DR EMBL: M37647; AAA26390.1; ALT_INT.
 DR EMBL: AF161079; AAD42234.1;
 DR EMBL: AJ235273; CAAL5140.1;
 DR InterPro: IPR003858; rompa_ompb.
 DR Pfam: PF02708; rompa_ompb; 1.
 KW Antigen; S-layer; Cell wall; Complete proteome.
 FT CHAIN 1 1328
 FT 1229 1643
 FT VARIANT 257 257 32 KDA BETA PEPTIDE.
 FT VARIANT 1010 1010 V -> A (IN STRAIN BREINL).
 FT VARIANT 1450 1450 Y -> D (IN STRAIN BREINL).
 FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).
 FT CONFLICT 178 179 AA -> VC (IN REF. 1).
 FT CONFLICT 191 201 TTQGAFLTGA -> INSRSSVHLVS (IN REF. 1).
 FT CONFLICT 212 212 T -> I (IN REF. 1).
 FT CONFLICT 313 313 Q -> L (IN REF. 1).
 FT CONFLICT 1104 1104 D -> G (IN REF. 2).
 FT CONFLICT 1123 1123 T -> S (IN REF. 2).
 FT SEQUENCE 1643 AA; 169854 MW; 735FDF392E6346CC CRC64;

Query Match 4.5%; Score 480; DB 1; Length 1643;
 Best local similarity 22.2%; Pred. No. 5.2e-11;
 Matches 396; Conservative 222; Mismatches 619; Indels 544; Gaps 97;
 QY 581 TNSKQIOVGADGTFITDISKPGATG-----NTTITTDGIGFANNESLDANKPRL 635
 DB 18 TASTATVAGSGVAM-----GAAMQYKRTTAAATTTDGIQFDQAG--ANTPVA 66
 QY 636 PTGGINAGKELTNQASINPAT-----NGOLDPMRLSTANKESSGAATINDLY 687

Db 67 PMSVITA-----NANPEITPTNPNGLNSLFLDTFANDLAVTINEDT---TLGFTT 113
 QY 688 NLSQVPLEFAGDTPGPNYTKLGEILKYVGKRTAD-----LTKNNIGVADSTD 737
 Db 114 NIAQAKFE-----NFTVAAGKILNITGGITVQEASNNIINQNALTKYHGGALINAND 167
 QY 738 ----NSLTVLAKLTLSDLDVANK-----TLASDKVTVDSGNNTAKLQNGDLTFESKON 787
 Db 168 LSGSGSTTFEAAAPSVLEFNNLNPPTOEARPLTLGANSKI-VNGGNGTININTEGFIQVS-DN 225
 QY 788 TGAATPNSKTIQVYDGLKFTDNNNGIALDGTYYITKQKVGFAKQDGLSKRPYLDKDKL 847
 Db 226 TFA-----GIKTINID-----DCQGLMFSNP-----DAANTL-----NMQ 256
 QY 848 VGEVEITTPNGINAGKAITGLSN---TLTATNATGHTVQLGIVDSTDKTAAASI---G 901
 Db 257 VGGNTINFNIGIDGTGLVIVSKNGAATEFNTVGLGSKNL--GIIELENTAAVAGKILISQ 314
 QY 902 DVLNAGNKLNN-GDADEVSTYDTYDFINGNATTAKVYDGAKSVAYDVNDG---TTI 958
 Db 315 GAANAIVICTNGAGRAAGFTVSYD---NGNAATISGQYAK-NMVIQSANAGQVTFE 368
 QY 959 HLTCAGDNKNQIGVYKTTTLTKRTAKGDKAINFVSNGDGLALINADIANMLTLAGEIR 1018
 Db 369 HIV-----DVLGTTNKT-----ADSKVITTENSNFSSTNF-----401
 QY 1019 NTKGTADTALOTFOVKKVEN---GDDDNADT-----ITWGD---AKTQNV 1061
 Db 402 ---GNLDTQIVPDTKILKNFTIGDYKNNGNTAGVITFMANGALVASSTDPNIAVFN-IN 457
 QY 1062 TTKLKG-----NGLDIOTNKGIVTFEGINTOSGLKAGNNTTLNNGLS 1105
 Db 458 AIEEGAGVVELSGIHIAELRLNGSGSIFKLADGTIVNG-----PVNOMALNNMNA 510
 QY 1106 IKNTAGMEQIYQVADGVKFAKVNNGVAGIDGTTTRITREIFAGTNGSL-----PK 1158
 Db 511 ---AGSIQ-----LDSALITGD-IGNGVNALOHITLANDA 544
 QY 1159 SKPHLSKDG-----INAGKRITNIOGSELAONSDAVTGKHYDTEU 1203
 Db 545 SK-LLALDGANIIGANVGAIHFQANGCTIKLNTQ-----NNIVNFDL-DITTK 594
 QY 1204 ENKISSAKTAONSLH-EFSVADQGNFTVSNPYSSTYDTSKTSVYITRAGENGITTKVN 1262
 Db 595 TGVVADSSLTNNOTLTINGSIGTVANTKTLAG-LNIGSSKT---ILNAGDAVIMELVI 649
 QY 1263 KGVAVRGIDOTKGLTTPKLTGVNNGKGIYIDSONCONGITTSNLTAVNTDKGSVRT 1322
 Db 650 ENNGSVQJLNNTYLTIKTITMAAN---QGOIYAADPLNTNT---TLADGTN-LGSA--- 698
 QY 1323 EOGKIITKDEKTRAASIVDLA--GFNLQNGEAVDFVSYD---TVNFADGNATTAK 1376
 Db 699 -EMPLSTIHFAKANAADSLILNKGCVNLVANN---ITTDANVAGSLHPSGGSISYS 752
 QY 1377 VNY--DDTSKTSKVYVDVAVDOTTIEVKKKIGVKTTLTSTGTGANKALSNQATGDL 1434
 Db 753 GTVGGGQGHNLNLLD---NGTIV---KFLGDTFNGCTKIEGKSIIQISNNYTTD-- 804
 QY 1435 VKASDIVAHLNLTSGDIQAKGASQANSAGY-----VDADGKRVYDSTDKKYGA 1486
 Db 805 VESADMTGTLFVNTPIV---TLNKGATFVGLVKQYIISGPGN-IYFNEI----- 852
 QY 1487 KNDGYDKTRKVAADKLVAAQOT-----PDGT-LAOMNVKSVINKEQVNDANKKGIN 1538
 Db 853 ---GNVGIYHGIANASISEFNASISLTLPLSGTPLDVLTIKSYGNGFVDN-----FN 903
 QY 1539 EDNAFVKGLEKASD-----NKTNAAYVVDLNAVAQPLPFPADDTGTAKKLGSETLTI 1593
 Db 904 APIVAVSGIDSMINQOIIIGDKNIIILISGDSNSTVNAVANTLYSGIRTKNNQO--TVTL 962
 QY 1594 KGGQ-----TDNKLTD-----NNI-----GVAGTD 1615
 Db 963 SSGMPNPNPGTIYGLGLENSPKLYQVTTFTDINNLSGIANNVTTINDYVTLITGGIAGTD 1022

QY 1616 GFTVKLAKDLTMINSVNAGCTKIDKGVSEVDS-----SGQAKNFPVLSANGLD 1665
 Db 1023 -PBAKI-----FLGSYNG-----NANRVFVSTSDPSMSIYVAIQANKGYTTY----- 1064
 QY 1666 LGGKVISNVGKGTQDDAANVOQLNEVRNLGLGNAGDNADNOVNIADIKKDPNSGSS 1725
 Db 1065 LCNALVSNL--GSLDTPVASVR-----FTGNDGSGAGLGONILYSQID---FGT 1107
 QY 1726 SNTVYKAGTVLGGK-----GNN-----DTEKATGV--QVG 1756
 Db 1108 YMLTILNSVVLGGGTFAINGEIDLTLNMLIPANGSTWGDNTSISTLNVSSGNGGVV 1167
 QY 1757 VKDG-----NANGDLSNVWVKTKQDSSKKALLATFYAA--GQTYNLNN 1799
 Db 1168 IAEADQVNAVTTGTTTITKIDQANANFSGTQAVTLIOG---ARENGTIGAPFEAVTG 1222
 QY 1800 PAEADIRINEGIRFEHVNDQNEPVOGRNGIDSSAGKHSVALGFOAKADGEAAVAG 1859
 Db 1223 -----SNIFVKYELLRDSNODVYLFRTNDVLNVVT-----1252
 QY 1860 KOTQAGNOSIALGDNAQATGDO--STAIGTVNVACKHSGALGDBS-----TYKADNSY 1911
 Db 1253 --TAVGNSAIA--NAPGVSONISRCLESTNTAAVNNMLAKDPDVAATFVGAIAITDTS 1307
 QY 1912 SVG-----NNOPTDATTQDVFVGNNITVYESNVALGSNSAISAGTHA-----GTQAK-- 1961
 Db 1368 AVTYVNLNDQTOQLDLSNRGLTLYLSNAETSDVAGATGAVSSGDAEVSYGVWAKPF 1367
 QY 1962 -----KSDGTAGTTTAGATGVYK--FAGQTAGAVSVGASGABRRIONVAAGEVS 2011
 Db 1366 YNIAEDOKKKGIAQ--YRAKTTGVVGLDTLASDMLIGAIIKTIDIKHQQDKKD-- 1423
 QY 2012 ATSTDAVNG--SOLKAKQOGLAN-----ATNELDHRHONENKNAOISSAMAMAS 2060
 Db 1424 --KTD-INGLSFSLGSOOLYKNEFPAQGNALFTLKKVKSQORFFESNGMSQOIAGN 1480
 QY 2061 MPQAYIPGRSMYTGIAITHNGGAVAY-----GLSKLSNGQ 2097
 Db 1481 YDNMTF--GONLIFG--YDYNAMPNVLVTPMAGLSYLSKSNE 1518
 RESULT 7
 FHAB BORPE STANDARD; PRT; 3591 AA.
 AC P1255;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Filamentous hemagglutinin.
 GN FHAB.
 OS Bordetella pertussis.
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 OC Bordetella.
 OX NCBI_TaxID=520;
 RN (1)
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=90355839; PubMed=2388559;
 RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
 RT "Genetic characterization of Bordetella pertussis filamentous
 haemagglutinin: a protein processed from an unusually large
 precursor."
 RL Mol. Microbiol. 4:787-800(1990).
 RN (2)
 RP SEQUENCE OF 1-3261 FROM N.A.
 RX MEDLINE=89202384; PubMed=2539596;
 RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
 RT Filamentous hemagglutinin of Bordetella pertussis: nucleotide
 sequence and crucial role in adherence."
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).
 CC -I- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND
 INFECTION.
 CC -I- SUBCELLULAR LOCATION: SURFACE.


```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M60351; AAA22974.1; -
DR EMBL; M60351; AAA22975.1; ALT_INIT.
DR EMBL; M60351; AAA22976.1; ALT_INIT.
KW Antigen; Hemagglutinin.
SQ SEQUENCE 3591 AA; 367420 MW; EF7418B30D6E5138 CAC64;

Query Match          4.58; Score 478.5; DB 1; Length 3591;
Best Local Similarity 19.94; Pred. No. 1.5e-10;
Matches 492; Conservative 306; Mismatches 900; Indels 769; Gaps 110;

OY 9 ENKATGTFMAVAEYAKSHSTGCG--GSCATGVGVSTL---SPRIALAVLVIGAT 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db YDHATRRATPIAAGARGAAGAYVIDGTAGAMYGKHTLVSSDSGLGVQLGSLSSPSA 323
OY 61 LNSGAVAGTISEADGGKGNANRGKSIATGDIQAALGSGSIATIGDNKIYHNSNNANI 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 324 ITVSSQGEIALGDATVVRGPISTLKAGAVSAGKLAGGGAANAVAGGAVKIASASSVNL 383
OY 121 GAKKSGNESIAI---GGDVLASGHASIAIGSDLLYKRETYOQISELPIIRGKALNDI 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 384 AVGGGKQVATLLNAGGTLVSGRAVOLGAS-----SRQLASVNA 426
OY 178 YQLADNLTQKRYRTHAGHASTAVGAMSYAKGHPNSAPGTA-----TAE 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 427 GALKADKLSATRRVAVGKQAVALSAS-----SNALSVAGGALKAGKLSATGRDVD 480
OY 223 GTSYLAAGLTATTAASIAVGSNAOA---IGFAATVAGSTOVNLTN---RG---IAL 271
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 481 GKQAVTLASVA-SDGALSVSAGNLRANETVSSKQLEVRGQREVALDASARAKMTYVAA 539
OY 272 GF-----GSOVLQKDNVNAANAVAYAPDDNPIIDNRYKATFKNGATVDSIGNSN 322
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 540 GALAARNLQSGALGVQCGEAVSANA-----NSDAELFRVG---RGVDLHDLASAR 589
OY 323 GNDSTRRKIIVGAGSADTD-----AVNVAQLKEAVRL---ANRQITFGDDSNRVE 372
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 590 GADISGEGRVNIIGARSRSDVKSVAHSAHGALSIDSMALGALGVQAGVSADKMSRGAV- 648
OY 373 KGLGKTLITGGAQTSALTIDNIGVYONGDLKYQLAETLTLKMWVTENTL---ANEKVT 430
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 649 -----TVSGGAV-----NLGDVQSDGOVRAITSAGAMTVRDVAAAADLALQGDALQ 695
OY 431 VGRTRLTDTKIGFTNDMGIDESKPYLDKDTGIIHAGGOKITKRLTAGVDDDAATYQOLK 490
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 696 AGFL-----KSAGAMTVGRDAVR--LD---GAHAGGQ-----LRVSSDQALGSLAA 739
OY 491 VNQTAESLQTFYKKVDKNGNDANDSKITTVGKNNKPDGQVQNTLKLKGENGVDTTET 550
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 740 KGELTVSAARAATV-----AELKSDINISVT--- 765
OY 551 NGTVFGLNONGTLVGNSTLNDGLSVKNTNSNKOIOVGADGITFDISKSGAGLEIN 610
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 766 -----GGENVSQVSNASRAVAISAHGALDGVKVSAGKSIGLEG 804
OY 611 TTRITRDGIFANNTGSLDANKPLRPTGINAGGKELTNVOSAINPATNGQLDFMRLS 670
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 809 WCAGVADSLG-----SDGAISSGRDAVRVQDA----- 832
OY 671 TANTKSSGAATIKDLVNLQVPLTFADTGPBNYTKKGLGELLKVGSGTTADDLTKNNIG 730
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 833 -----RSLDISLGAEGATLGAVEAAGSI-DVNGSGTVA----- 866
OY 731 VVADSTNSLVKLAKTISLDLDAVNTKTLTASDKVTVDSGNNTAKLQNGDLTFKQNTGA 790
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 867 -----ANSLHANRQVRSXGDAVRVTAATSGGGLHVSSGR-----OLDIGA 907
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 791 LPATNSKTIGVDGLKATDNNNGIALDSTYITTDKGFAGQDQSLD-KSPFYLDKRL-KV 848
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 908 VQARCA--LALDG-----GAGVALOS-----AKASGTLVHOGGEHLDTGLAAV 949
OY 849 GEVEITNGINGKKAITGISTNLTPDA--TNATGHTVTOGIYDSHPDKPRASIGVLNAG 907
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 950 GAVDV-----NGIGDVRYAKIVSDAGADLQAGRSMTLGIYDTTGDDLQARA-QQKLEIG 1001
OY 908 FNLKNGDAKDFVSTYDVTDFINGNATTAATYVYDGKASVAVDVNDGTTILHTGADNGK 967
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1002 -SVKSDG-----GLQAAAGALSLAAAVAGALELISGGGVYVDRBASAR 1044
OY 968 -----NOIGVKTTLTKTDA---KGDAL-----NFSVSGD-DKAL 1000
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1045 ARIDSTGVGICALKAGAVEASPRRARALRODFTPGSVVAVAGNTVVRGDPHOGV 1104
OY 1001 INKDLADNLNTLAGELRNKTKGADTALQTFQYKKYKENGDDNDADITTVGKDAKTNOV 1060
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1105 LAQGD-----IMAKG--GTL--LRNDALTEMG-----YVTSADSAVLEH 1143
OY 1061 NTKL-----KKNGLDIQTNK---DGTVEGINTOSGLKAGNNTLNNGL 1104
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1144 STIESKISQSVLAKADKDKGPRASVAVAKKFLNGT-----LRAVN---DNN-- 1187
OY 1105 SIKNTAGNEQIOVGADVVFAYKAVNGVAGADGTTTRITRDEIGFAGTNGSLDKSPHUS 1164
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1188 -ETMSGRIDV-----VDGRPOIT-----DAVGEARKESVS 1220
OY 1165 KDGINMGKKTITNIOGGEIAONSNDNAVGTGKIYDALTLELNKISSAKTAONSLEH-FSV 1223
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1221 DAALVADGPTV-VEAGELVSHAGTONG-----RNKENGASVYRTTGLVYNGYIS 1272
OY 1224 ADEOG-----NNFTV--SNPYSSYDTSKTSYDTF-----AGENG-ITTKVNGVVR 1267
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1273 AGROGVLENGALNLEFLVSGDGTORIEAQRIENRGFPQOAPAGTAGALVVAABAIYH 1332
OY 1268 VGDQTKGLTTPKLYVGNNGKGIIVDSQNGQNTITGLSITLANVTN-----DKSVPT 1321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1333 DGWATKGG---EMQIAGKGGSPPTVYAGAKATTSANKLSDVASWPNAGSLDKIKGGAOY 1389
OY 1322 T-----EOGKITKDEKTRTAASIVDLVSAGFNLOGGEAVDVSYDTF----- 1364
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1390 IVAGRYAHEGEVSIQGYTISA---DAIALAQYTQRGGAANLTSNDRIFSKIRLMGP 1446
OY 1365 ---VNFADGNATTAATKYDDTSKTSKVYVDVNDVDTTIEVKDKLGVKTTLLSTGTGAN- 1421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1447 LOVNAGGPPVNTGMLKRVREGVYTAASF--NETGAEVN-----AKSATLITSGAARNA 1498
OY 1422 -KPAISQATGDLVAKASDIYVAHLNLTSGDIOYAKGASQANSAGVVDAG-----AKVIY 1476
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1499 GKMQVKEAAT---IVAAS--VSNGGFTTA-----GNDITVTSRGGEDMEGKMSKNDIV 1547
OY 1477 DSTD-----NKYYQAKNNGYVDKTEV-----AKDLVAQAQTPD--GTL-----AOMN 1518
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1548 IKTEQPSNGRYLDKHLDTFYASQADNRGSLAKAGHDPTVOAQRIDNSGTMAGGHATLK 1607
OY 1519 VKSVYINKQVNDANKKOGIEMDNAFVGLBEKASDNKTKNAAYT-VGLNVAVAQPLPTEA 1577
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1608 APLHRTNGVYAGHDHIIINSKLENTGVRDARDLADLVDADFTNGSILYAEHADTLTLA 1667
OY 1578 GDT-----GT-----TAKLGEITLIKGO-----TDNKLTLDNNGIVAG 1613
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1668 QGTQRDVLVVDHLLPVAEGTLRVKAKSL--TTEIEFGNGLSLAEVQENIDMKQAVIG 1725
OY 1614 TD-----GFTVKLAKDLTINLS--VNAAGTKPIDDKGVSEFVDS 1649
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1726 KDLTLSSAHGNVANEANALLMAGELTVK-AQNTTKRALIFRGGARALTAVALLNKL 1784
OY 1650 GQAKANTPVL SANGLDVGKAVISVNGKGTQDPAANYQOL-NEVRNLGLG-----NA 1701
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Db 1785 GRIRA-----GDMHLDAPRIENTAKLSGEVQRKGVQDVGGENGHRMSGIGYVNWMLRA 1838
 Qy 1702 GNDNADNOVNINDIKKDNSSG--SSNRTVIKAGTVL-----GKGNNTEKLTAGGVQV 1755
 Db 1839 GNRKRG---TLA-----AFWYGGDLTAEOSLIEVGKDLTLNAGARDEHRHLLNEGVIA 1891
 Qy 1756 GVDKDNANGDLSN----- 1769
 Db 1892 G--GHGIGGDVDNRSVFTVSAMEYFKTLPVSLTALDNRAGLSPATWNPSTYELLDY 1949
 Qy 1770 -----VW-----VKTKD----- 1777
 Db 1950 LLDONREYIWLGYPTYTEMSVNTLKLNDLGOAKRPAAPPMPKAPBELDRLGHTLESAN 2009
 Qy 1778 -----GSKKALLATYNAAGOTVTLNPPAIDRI NEO--GIRPFYHNDGNGPPVQGRNG 1831
 Db 2010 GKRIEYKLLQGEYKAKAMAVQAVEATREYRHVDLQGRYKALGSMDETVEVDGI 2069
 Qy 1832 IDSSASGKHSVAI--GFOAKADGEA-----AVAIGROTAQNGOSIAIG 1872
 Db 2070 IGEFADLRTVYAKQADQATIDETDKVAQRYSQDIDAVRLQAIQPGRYTLAKALSAALG 2129
 Qy 1873 DNAQATGDSIAIGTGNVYAGKHSAGIDPSTYKADNSTSVGNNOFTDATGTDFVGCYN 1932
 Db 2130 ADMRALGHSOLQORWKDFKAGRGAEIA-----FYPRKQT--VLAAGA 2170
 Qy 1933 NITVESNSVALGNSAISAGTHAGTO--AKKSDGAGTATTAGATGVYKAGQOTAVGA 1990
 Db 2171 GLTIS--NGAIHNGENMAONRGPRPEGKIGAHSAITSVSGFDLRVGLKRLDIDDALAA 2229
 Qy 1991 VSYG-----ASGERRIONVAAGEVSAT-----STDAVNGSOLYKA 2026
 Db 2230 VLVNPHIETRIGAAQOTSLADGAAGAPALARQARQAPETDGMVDARGISDALASLASIDA 2289
 Qy 2027 TQGINATNELHRIHONENKANAGIS--SAMAMASMPQATIPGSMYTGATITNGGA 2084
 Db 2290 AAGL-----EVSGR--RNAQVADAGLAGPSAVALPAGVADV--GVEPTVG--DQVDPV 2338
 Qy 2085 VAVGLSK 2091
 Db 2339 VAVGLEQ 2345
 RESULT 8
 HLYA_PROMI STANDARD: PRt: 1577 AA.
 ID HLYA_PROMI STANDARD: PRt: 1577 AA.
 AC P16466:
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-NOV-1990 (Rel. 16, Last annotation update)
 DE Hemolysin precursor.
 GN HPMA.
 OS Proteus mirabilis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Proteus.
 OX NCBI_TaxID=584;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-43.
 RC STRAIN-ISOLATE 477-12;
 RX MEDLINE=90170827; PubMed=2407716;
 RA uphoff T.S., Welch R.A.;
 RT "Nucleotide sequencing of the Proteus mirabilis calcium-independent
 hemolysin genes (hpmA and hpmB) reveals sequence similarity with the
 RT Serratia marcescens hemolysin genes (shlA and shlB).";
 RL J. Bacteriol. 172:1206-1216(1990).
 CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
 CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
 CC DEFINED.
 CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
 CC ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. HPMA
 CC REQUIRES HPMB FUNCTION.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- MISCELLANEOUS: THE CONSERVED AMPHIPATHIC DOMAINS IN SHLA AND HPMA

CC MAY BE RESPONSIBLE FOR PORE FORMATION.
 CC -!- SIMILARITY: TO S. MARCESCENS HEMOLYSIN (SHLA)
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; M30186; AAA25657.1; -
 CC DR PIR; A35140; A35140.
 CC KW Hemolysis; Toxin; Outer membrane; signal.
 CC FT SIGNAL 1 29
 CC CHAIN 30 1577 HEMOLYSIN
 CC SEQUENCE 1577 AA; 165869 MW; 175975E0C924B2D9 CRC64;
 SO
 Query Match 4.48; Score 474.5; DB 1; Length 1577;
 Best Local Similarity 21.88; Pred. No. 8e-11;
 Matches 397; Conservative 209; Mismatches 597; Indels 615; Gaps 96;
 Qy 506 KVDKNGDANDSKITTYGKNNKPDGTQVNTLKLKGNGVDVTTETNGTVFGL----- 558
 Db 7 KLSPSGRLASLALIFVSLNAYGNGIVDA-----GHGQPVSAVANGSTQVINITVTPNNEG 62
 Qy 559 ---NQ-----NGLVYGNSTLNNDGLSVKNTSNKQIQVYADGCTFTDISNS 602
 Db 63 ISHNQYDFENVKPGAVFNALFAGOSQL-----AGHLNANSNLNGQAASLILNEVSR 116
 Qy 603 KP-----GAGIE-----NTTRIRDIGFANNNGSLDANKPRLTPTGIANAGKELT 648
 Db 117 NPSFLDQGEVFGIAAEVYLSNPGITCDGCFINTSRS-----SLVVG----- 160
 Qy 649 NVQSAINPATNGQLDFPMNRLSTANTEKSSAATIKDLYNLSOVPLTFAGDTGPNYTKKL 708
 Db 161 -----NPLEFNGQLKGYSTLNTN-----LLSLCKNGLNTTGLL 194
 Qy 709 GEILK--VKGKTTADDLTKNNIGVAVDSTDNLSITVYLAFTLSDLDVAVT--KLTLSDKV 765
 Db 195 DLAPRIDSRSKTIYAEIS-----AFTGNTFS-----OHFDLSQKPSALDSY 240
 Qy 766 TVDSGNNTAKLQNGDLTFESKONTGATPATNSKITGVGLKFTDNNGLALDGTYYITKDY 825
 Db 241 EFGS-----MQSRIITITAE--SGVKLAG--KFTINDLSYKADINQDSQV 286
 Qy 826 GFAKQDSLKSKRPYLDKRLKVEVEITTINGINAGKATIGLNTLTATNATGHWYQ 885
 Db 287 RYDSYD--KDGSENY--QVVRG--ITVNN-----SGSSQTLT--KTELKGNITL 329
 Qy 866 LGYDSTDKTRASI--GDVLNAGFNILKNNGDAKDPSTVDTYDFINGNATTKAVT----- 939
 Db 330 --VASSHNQIKASDLMDGDDTLTGADLT-----IDGQLOLOKEDIDN 370
 Qy 940 ---YDGKASVAVYDVNVDGTTIHLTGADGNKNOIGVY--TTLTKTKDAGKRAINFVS 993
 Db 371 RWFESMK---YDVYKEKQIQIG-----SQIDAKNNATLTATGDTLDA--KTN 417
 Qy 994 SGDDKALINKADIADNINLTAGEIRNTKGTADTALQTFQYKVKENGDDNDADITTVK 1053
 Db 418 AGNNLAINAKDI--HINGLV-----EKESRSSENGKRNHRSLESGS 458
 Qy 1054 DAKNQNVTKLK--GKN-GLDIQ-----TND 1078
 Db 459 WSNHQETLKASBLTAGKGLDAGOSTTAQCAKLHANEVLNVAKNININLVQKTND 518
 Qy 1079 GTVFFGINTQGLKAGNNTLNNGLSIKNT-----AGNQIOVAGDGVY----- 1123
 Db 519 KTYIDNRYMMGGIGGQGNK--NNNQOVSHARPULTADGOLLAAADNNVITSGVYKNG 577
 Qy 1124 -PAKVNNGV-----GAGIDSTRTIDDEIGFAGTNGS---LDSKSPH 1162

| | | | |
|----|------|--|------|
| Db | 578 | AFVYTTQGDVADNALSETISKIDERTGFAFNITSSKHNNENNQTSGSELISDAQTLV | 637 |
| Qy | 1163 | LSKQCIAGGKKTITNIO\$-----GELAONSDAYTGKITYLQKLEENKISSRAKTAQN | 1216 |
| Db | 638 | VSGDNVAVVLSIK\$ADKLGLHSLGJINX\$AOQVY- KIDDEKTSL- AITGAK- --- | 689 |
| Qy | 1217 | SLHFSVADQ\$---GNNEF- VSNP\$SYDP\$KTSQDVTFPAGE\$GTLTFVYNNGVAVYDQ | 1272 |
| Db | 690 | -----EVEDKQYSAGFHITHTTKNTSTEBQAN\$---TISGAN- VDLQNNKDVTFA\$GDL | 741 |
| Qy | 1273 | TKGLTTFKLTIVYNNN- GKGIVY- -----DSQNGONTITG----- | 1304 |
| Db | 742 | -----KTAGNAN\$ITGDVNAFVSTENKQKQIDNDITLISGFYSYTGVDYK\$ADPQY | 794 |
| Qy | 1305 | -LSMTLANVTNDKGSVTEBQKI\$KDEKTRASIVDLSAGFN- -----LOGNGEAYD | 1357 |
| Db | 795 | DKOHTQEVTEKNR\$S- QTEVAGD\$LITLTKAN\$-----DLH\$GASHHVEGRQ\$E\$GENIQ | 846 |
| Qy | 1358 | FV\$TYDYNFADGNATTA\$KTYTDTSKTS- -VYVDVNDQDITL\$EYDKL\$---G\$VKT | 1410 |
| Db | 847 | HLAAND\$-----ET\$KTSISL\$NYDGVNLDYSGVTP\$KKAIEDGNT | 890 |
| Qy | 1411 | TLT\$TGANKFAL\$NOATGDALV\$KASD\$IV\$HLNTLS-----GDITQAK | 1455 |
| Db | 891 | T-----KPCNN\$D\$LT\$KVI-----ARDAI\$NL\$ML\$NLETP\$NV\$EVEG\$IKG\$SSQ\$SQ | 938 |
| Qy | 1456 | GASQAN\$---SAGYVDADGNKVIYD\$TDNKKYQAKNDG\$VTDKTEV\$AKDKLVAOQ\$PDG | 1512 |
| Db | 939 | TDSQAV\$T\$INA\$KIDID\$SN\$K\$LHD- QGTHYQ\$TQEGI-----SLTANHTSEA | 986 |
| Qy | 1513 | TL-----ADMN\$K\$V\$YNKEQ\$V\$V\$D\$ANK\$OG\$INEDNA\$F\$Y\$K\$GLEK\$AS\$DMK\$TNA | 1555 |
| Db | 987 | TLDKHQ\$TTFHETKGGQIGV\$T\$KTSQD- IYVAIK\$GQ\$T\$D\$ALMETK\$K\$GQ\$T\$SNGDI | 1045 |
| Qy | 1560 | AYVYD\$-----LNAVAQ\$PL\$F\$AD\$T-----G\$T\$AKLEETLTI | 1593 |
| Db | 1046 | SINNGENAHYGAOP\$D\$AKQ\$K\$YV\$IN\$AGD\$ITL\$Q\$AD\$H\$T\$SE\$OS\$V\$N\$G\$AN\$K\$VGT\$PES | 1109 |
| Qy | 1594 | K-----GQOTD\$YNK- L\$D\$NNIG\$V\$VA\$D\$GT\$F\$V\$K\$LA\$K\$DT-----N\$N\$SYNA | 1633 |
| Db | 1106 | KDYGGF\$NA\$G\$T\$H\$H\$K\$EQ\$T\$AK\$VGT\$IT\$G\$SG\$IEL\$NG\$H\$N\$LT\$LOG\$H\$L\$SE\$OD\$IAL\$N\$FV- | 1164 |
| Qy | 1634 | GGT\$K\$IDDK\$Y\$F\$V\$D\$SSQ\$AK\$AN\$F\$V\$LS\$AN\$GLD\$G\$K\$V\$IS\$N\$G\$K\$GT\$D\$D\$A\$N\$Y\$Q\$LN\$EYR | 1693 |
| Db | 1165 | ---KVDLOSAS- SEHTEKGN-----NL\$SGVQ\$A\$F\$G\$K\$K\$MTD- DASSVNG\$IG\$SQ | 1206 |
| Qy | 1694 | NLLG\$-----L\$G\$AN\$GD\$NAD\$GN-----Q\$V\$N\$D\$IK\$K\$D\$P\$N\$G\$---SS\$NNT | 1722 |
| Db | 1209 | FALIGQ\$DEK\$YS\$R\$B\$G\$T\$IN\$N\$G\$LT\$IN\$G\$N\$SV\$H\$Q\$A\$V\$N\$S\$K\$D\$Q\$LT\$S\$G\$S\$D\$IELT\$Y\$A\$OST | 1266 |
| Qy | 1730 | V\$IK\$A- G\$TVL\$G- G\$K\$GN\$D\$T-----E\$K\$LAT-----G\$VQ\$V\$G\$V\$D-----K\$D\$G\$AN\$G\$D\$S | 1766 |
| Db | 1269 | DYKNNW\$G\$D\$IG\$FN\$K\$K\$T\$IN\$TP\$K\$E\$T\$E\$K\$P\$AT\$S\$IH\$N\$IG\$K\$LL\$V\$N\$E\$D\$Q\$K\$T\$SH\$Q\$N\$LT\$E\$G | 1322 |
| Qy | 1769 | NW\$V\$K\$TQ\$D\$---G\$SK\$ALL\$AT\$Y\$N\$A\$G\$Q\$N\$Y\$V\$LT\$N\$P\$A\$E\$ID\$R\$IN\$E\$Q\$IR- F\$F\$HY\$N\$D\$N | 1823 |
| Db | 1329 | TLTINS\$K\$D\$LT\$G\$AN\$V\$Y\$AD\$SV\$T\$G\$N\$G\$S\$LIN\$A\$Q\$KES- -D\$R\$H\$Y\$V\$G\$V\$N\$G\$H\$N\$D\$PK | 1384 |
| Qy | 1822 | Q\$P\$V\$VQ\$G\$N\$G\$D\$S\$AS\$G- -K\$H\$V\$AL\$FQ\$AK\$AD\$G\$E\$A\$V\$AL\$IG\$Q\$T\$O\$AG\$N\$ST\$A- --- | 1876 |
| Db | 1386 | SSQ\$V\$N\$K- -T\$A\$R\$G\$S\$LE\$K\$T\$IK\$D\$IT\$D\$G\$IS\$S\$TD- -A\$ID\$K\$Y\$S\$LS\$T\$AD\$K\$TG | 1433 |
| Qy | 1871 | ISD\$N\$Q\$AT\$G\$D\$S\$IA- I\$G\$T\$G- N\$V\$A\$G\$H\$G\$AL\$IG\$D\$P\$TV\$K\$AD\$N\$G\$S\$V\$G\$N\$N\$Q\$F\$D\$AT\$Q\$TD\$V | 1927 |
| Db | 1436 | ISD\$E\$K\$AK\$ID\$G\$F\$G\$K\$V\$G\$IN\$G\$IN\$Y\$T\$G- -A\$B\$G\$H\$A\$M\$D\$IK\$V\$TH\$V\$D\$N\$D\$A\$Y\$K\$T\$- --- | 1488 |
| Qy | 1928 | F\$G\$Y\$N\$N\$IT\$V\$E\$S\$N\$Y\$AL\$G\$N\$S\$A\$IS\$AG\$H\$A\$G\$T\$Q\$AK\$S\$D\$G\$T\$T\$A\$G\$AT\$G\$Y\$K\$G\$F\$AG\$Q\$TA | 1987 |
| Db | 1486 | -----S\$LT\$S\$N\$N\$D\$SL\$V\$N\$-----G\$ST\$K\$IG\$A- E\$IV\$S\$Q\$G\$O\$V\$D | 1514 |
| Qy | 1988 | V\$G\$A\$V\$S\$Y\$G\$AS\$G\$A\$R\$R\$IQ\$N\$Y | 2005 |
| Db | 1517 | LG\$G\$S\$V\$-----K\$LE\$N\$T | 1527 |

| ID | OMPb_RICCN | STANDARD | PRF | 1655 AA |
|----|--|-----------------------------------|-----|---------|
| AC | 09KKA3 | Q9KX98; Q9XC45 | | |
| DT | 16-OCT-2001 | (Rel. 40, Created) | | |
| DT | 16-OCT-2001 | (Rel. 40, Last sequence update) | | |
| DT | 01-MAR-2002 | (Rel. 41, Last annotation update) | | |
| DE | Outer membrane protein B precursor (168 kDa surface-layer protein) | | | |
| DE | (Surface protein antigen) (Cell surface antigen 5) (Scat) (OMPb) | | | |
| DE | (comp.B) [Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide]. | | | |
| GN | OMPb OR R1085. | | | |
| OS | Rickettsia conorii. | | | |
| OC | Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; | | | |
| OC | Rickettsiaceae; Rickettsiae; Rickettsia. | | | |
| CC | NCBI_TaxID=781; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=Malish 7; | | | |
| RX | MEDLINE=21452074; PubMed=11557893; | | | |
| RA | Ogata H., Audic S., Remesto-Audiffren P., Fournier P.-E., Barbe V., | | | |
| RA | Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M., | | | |
| RA | Raoult D. | | | |
| RT | "Mechanisms of evolution in Rickettsia conorii and R. prowazekii"; | | | |
| RL | Science 293:2093-2098(2001). | | | |
| RC | [2] | | | |
| RP | SEQUENCE OF 33-1649 FROM N.A. | | | |
| RC | STRAIN=Indian tick typhus, and Malish 7; | | | |
| RX | MEDLINE=20393643; PubMed=10939649; | | | |
| RA | Roux V., Raoult D. | | | |
| RT | "Phylogenetic analysis of members of the genus Rickettsia using the | | | |
| RT | gene coding the outer-membrane protein (ompb) ." | | | |
| RL | Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000). | | | |
| RN | [3] | | | |
| RP | SEQUENCE OF 353-1655 FROM N.A. | | | |
| RC | STRAIN=Malish 7; | | | |
| RA | Stenos J., Walker D. | | | |
| RT | "The rickettsial outer membrane protein A and B genes of Rickettsia | | | |
| RT | australis, the most divergent rickettsia of the spotted fever group." | | | |
| RL | Submitted (May-1999) to the EMBL/Genbank/DBJ databases. | | | |
| CC | -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR | | | |
| CC | STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL | | | |
| CC | VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY | | | |
| CC | SIMILARITY). | | | |
| CC | -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR | | | |
| CC | (BY SIMILARITY). | | | |
| CC | -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S- | | | |
| CC | LAYER WITH HEXAGONAL SYMMETRY (BY SIMILARITY). | | | |
| CC | -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPb/OMPb FAMILY. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the Eutropen Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (see http://www.isb-slb.ch/announce/ | | | |
| CC | or send an email to license@isb-slb.ch). | | | |
| CC | ----- | | | |
| DR | EMBL; AE008659; AAL03623.1; - | | | |
| DR | EMBL; AF123721; AAF34124.1; - | | | |
| DR | EMBL; AF123726; AAF34129.1; - | | | |
| DR | EMBL; AF149110; AAD39533.1; - | | | |
| DR | InterPro, IPR003858; compA_OMPb. | | | |
| KW | Pfam, PF02708; compA_OMPb; 1. | | | |
| FT | CHAIN: 1 1334 | | | |
| FT | CHAIN: 1335 1655 | | | |
| FT | VARIANT 61 61 | | | |
| FT | VARIANT 75 75 | | | |
| FT | VARIANT 78 78 | | | |
| FT | VARIANT 251 251 | | | |
| FT | Variant | | | |

FT VARIANT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).
 FT CONFLICT 353 354 K -> G (IN REF. 3).
 FT CONFLICT 776 776 F -> S (IN REF. 3).
 FT CONFLICT 1159 1159 E -> D (IN REF. 3).
 FT CONFLICT 1177 1177 G -> S (IN REF. 3).
 FT CONFLICT 1492 1492 H -> R (IN REF. 3).
 SO SEQUENCE 1655 AA; 168342 MW; E49E19377D5FCE37 CAC64;

Query Match 4.3%; Score 456.5; DB 1; Length 1655;
 Best Local Similarity 22.0%; Pred. No. 4e-10;
 Matches 408; Conservative 191; Mismatches 677; Indels 579; Gaps 96;

QY 473 LTACVVDAAAT-----YGLKRVNOTAESALOTFFVKKVKKGDAND----SKT 519
 DB 12 ISAGLVASTATTIVASFAGSAMGAIAIOQNRITNAVATTVGVGFDQTAVPANAVAPLANAV 71
 QY 520 ITVGN-----NRPDGTQVNTLKLKGENGVDTETNGVTYFGLNONGGLTVGNSTLND 574
 DB 72 ITAGVNGGIILNTPAGS-FNGLEPLTANNIDYVREDTTLGFTTVVNNANHEFLMLD--- 127
 QY 575 GLSVKNTNSKQIQVADGITTFTDISKSPAGIENTTRITRDGIGFANNTGSLDANKPR 634
 DB 128 -----NACKTILITGOGIT-----NVOAATKKNANNV-----VAQVNGGAIDNND- 168
 QY 635 LTPYG-INAGKELTYOASAINPATNGQIDFENRLESTANTEKSGSAATIKDLYNLSQVP 693
 DB 169 LQVGRIDCGAASTLVFNLANPT-----QKAP 197
 QY 694 LTFAGDGPVNTKLGIEILKYKGGKTTADDLTKNNGVADS--TDSNLTVLAKTLL- 748
 DB 198 LI-----LGNAYIVGANGTLANVTNGFIKVSKEPATYVNIIGGQIGIMEN 245
 QY 749 SLDVAVTKTLASDKVTVDSGNNTAKLQNGDLTFKQNTGATPATNSKTIGVGLKFTD 808
 DB 246 TDADVNTMLQANGATITFENGTD---GTGRVLKSKNAATIDFVNTGSLG----- 293
 QY 809 NNGIALDGTIYITKDKYKFAKQSGSLDKSRPYDKDKLKEVEIITNGINAGKAITGL 868
 DB 294 -----GNLK-----GIEFTYVAVNGOLKANAG- 316
 QY 869 SNTLTDAITNATTHVTOLGIVDSTDKTRAASI-GDVLNAGFNLNKNGADPFSTVDYD 927
 DB 317 ANNAVIGTNNAGBRAA--GFVSVVDNGKATIDGOYV-----AKDW- 356
 QY 928 FINGNATTAKYVDGKASKVAYVNVDTTILHFGADGNKNOIGVKTTLTKTDAKDKA 987
 DB 357 -IGSANAVGVNF-----RHIVDVGDTGTAFKTA- 390
 QY 988 INFSVNG--DOKALINAKDIDLNLGAIKRTKGTADTALOTPOVKYKENG---DDD 1043
 DB 391 ITONSNEGTIDFENLAAQIIVPTMTLNG---NFTGDASNPGNTAGVITFDAGTLASAS 447
 QY 1044 NDADITTVGKADAKTNOYNTLKLKGNGLDIQUTKDGSTVFEGITQSGLKAGNNTTLNNG 1103
 DB 448 ADANV-----AVTNNTITAEASAGAVYOLSTHAEFLMG-NAGSEFLADGTIVNGK- 499
 QY 1104 LSKTNAGNEQIOYVADGVKFAKVNNGVAG--IDGTTTRITRDEIGFAGTNGSLDKSK 1160
 DB 500 -----VNOTALVG-----GALAGATTLDGSAITGD-IGNAGGAALOGIT 540
 QY 1161 PHLSDG-----INAGKKTINISGELTAONSNDNAVGTIKYDKLTELEKISTAKTAONS 1217
 DB 541 --LANDATYTLTGANITIGANGT-----NFQANGT-----KLTST-----QNN 581
 QY 1218 L--HEFSVADEQGNFTVSNPYSDTSTKSDVTFPAGENGITTFKNGGVAVGIDOK 1274
 DB 582 IYVDFLALATDQ---TGVDASSLTNAQT---LTKNKIGTVGANKTKLQDFINIGSSK 634
 QY 1275 GLTTP-----KLTVGNNN-----GKG-IVIDSQNGQNTITGLSNTL 1309

DB 635 TVLSDQVAINELVYINNNANVOFAHNTYLLITRTNAAGGKILFNPVYNNNTTLATGTL 694
 QY 1310 ANVTNDKGSVPTTEQKIKIDEDKTRAASIVDLVLSAGFNLOGNRAVPSYTDV----- 1365
 DB 695 GSATNPLAEINSGSK-----AANVDIVLANGVK- NLVATNTTTTDAVGSRI 742
 QY 1306 ANADGNATTAKYTYDTSKTSKVYDVAVDOTTIEVKDKKIGVY-----TTLSTGT-- 1418
 DB 743 ENAGGTIVSGVGGQKGFNTVALD---NGTV-----KFGNATFNGNTTIANSTIQ 795
 QY 1419 -GAN--KFLASQATGDALVKAASDVAHLNLTSGDIOTAKASQANSAGY-VADQNK 1473
 DB 796 ICGNTYADVASADTG-----IVEVNTGPITVILNKAAVYVNAKQITVSGPBNV 847
 QY 1474 VYDSTDNKYQAKN-DGTVDKTEKVAKDLVAQATP-----DGLAOMVKSVALNK 1525
 DB 848 VI-----NEIGNAGNVHGAVTDTIAFENSSLGAVFELPGIFENDAGRIPLTIKSTVGN 902
 QY 1526 EOVNANKKQGINEDNAFYKLEKASDNKT--KNAAVTVG--DLNAVAOTPLTFAGD 1579
 DB 903 KTAAT-----GFDVBSVIVLGVDSVYIADGQVIGDQNNITVGLGSDNDIIVNATTLVAG- 955
 QY 1580 FGTAKLGETTLIKGQTD-----NKLPD--NNGVAVGT----- 1614
 DB 956 IGTINNNG--TYTLSSGIPNTPGYVGLGTGIGASKFKQVFTTIDYNNLGNITATNATIN 1014
 QY 1615 DGEYVLA-----KDLTNLSVNAAGTKIDDKGVSEVDS-----SQOANAPVVL 1659
 DB 1015 DGVTVTTGIGIAGIDFGDKITLISVNGN-----NVRRVYDGLSHSTSMIGTANNQTV 1068
 QY 1660 SANGDLOGKVISNNGKGFQKOPDANVOOLNEVRMLLGLAGNNDNADNGOVNADIKD 1719
 DB 1069 TY-----LGNAPFNGI--GSDTPVASVR-----FTSGSGAGLOGINTYSOVID 1110
 QY 1720 ---PNSGSSNRFVIKAG--TVLGGKNDTEKLTAGVOVQVDRKGNAGDLSNVWVYQ 1775
 DB 1111 FGTYNLGSNSVILLGGGTALNGKINLRNTLTP-----ASG--TSTW----- 1152
 QY 1776 KQSGKALLATYNAAGOTNYLTNPAEALIDRINEGIRFPHVNDGNDEPVQGRNGIDSS 1835
 DB 1153 --GNNTSIETTLTLA-----NGNIGNIV----- 1173
 QY 1836 ASGKSHVAIGFOAKADGEAAVAIGROQAGNSTAIGDMAA--TGQSLAIGTGVVAG 1893
 DB 1174 -----IEGQAVNA-----TTGTITIKVQDNANANPSGTQTYTLIG---GA 1213
 QY 1894 KHSAGIDPSTV-----KADNSYVGNNOQFDTATQTDV-----GVG 1931
 DB 1214 RENGTLGGNPFVYVTSNRFVNVGLIRANQDYVITRTNANENVYTNDIANSFGCAGVYG 1273
 QY 1932 NNIT--VIESNSVA-----LSNSNAISAG-----THAGTOAKKSDGTACTTT 1971
 DB 1274 QWTFEVNATYTAANNLLKANSANSANFVAGIYVDTSAITNAOLDVAK-DIOADLGN 1332
 QY 1972 TAGA-----TGTVKGFACOTAVGAVSVGASGAERIORVAVG-BVASTSDA-----VNGS 2021
 DB 1333 RIGALRYLGTPEPAMAPPEA-GALPAVVAAGDERVAVAGIWAAPPYTDHAOSKKGL 1391
 QY 2022 QLYKA--TOGINAATNELDRITHONENKANAGISSAMASMP---QAYIPGRSMTVGTG-- 2075
 DB 1392 AGYKAKTGVVIGLDTL-----ANDNLM-----IGAIGITTKIDKHQYKKGDKTDVNGFS 1443
 QY 2076 -----LATHNGGAVANGSLKSLSDNGOVWRKINGSADTQGVGAAGAFH 2121
 DB 1444 FSLYGAQOVLVKNFPAQSAIFSLNQVKNKSQRYF-----FDANGNMKQIAAGHY 1493

RESULT: 10
 P3P_LACLC STANDARD; PRT; 1902 AA.
 ID P3P_LACLC
 AC P15292;
 DT 01-Apr-1990 (Rel. 14, Created)

01-APR-1990 (Rel. 14, last sequence update)
 16-OCT-2001 (Rel. 40, last annotation update)
 Pili-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-
 associated serine proteinase).
 GN PRP.
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
 CC plasmid.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 CC Lactococcus.
 CC NCBI_TaxID=1359;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
 RC STRAIN-SK11;
 RX MEDLINE=89340435; PubMed=2760036;
 RA Vos P., Simons G., Siezen R.J., de Vos W.M.;
 RT "Primary structure and organization of the gene for a procaryotic,
 cell envelope-located serine proteinase.";
 RL J. Biol. Chem. 264:13579-13585(1989).
 CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
 GROWTH OF THE BACTERIA ON MILK.
 CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
 specificity, although some substrate preference have been noted,
 e.g. large hydrophobic residues in the P1 and P4 positions, and
 Pro in the P2 position. Best known for its action on caseins,
 although it has been shown to hydrolyse hemoglobin and oxidized
 insulin B-chain.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 SUBTILASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL J04962; AAA03533.1; ALT_SEQ.
 DR PIR: A32634; A32634.
 DR HSSP: P00782; 2SPT.
 DR MEROPS: S08.019; -;
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR003137; PA.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR Pfam: PF02225; PA; 1.
 DR Pfam: PF00082; Peptidase_S8; 3.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
 DR Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
 KW Transmembrane.
 FT SIGNAL 1 33
 FT PROPEP 34 187
 FT CHAIN 188 1902
 FT DOMAIN 188 1876
 FT TRANSMEM 1877 1895
 FT DOMAIN 1896 1902
 FT ACT_SITE 217 217
 FT ACT_SITE 281 281
 FT ACT_SITE 620 620
 FT DOMAIN 1867 1872
 FT SEQUENCE 1902 AA; 200550 MW; 87CEBBA9345F9D3 CRC64;
 Query Match 4.2%; Score 444.5; Db 1; Length 1902;
 Best Local Similarity 20.8%; Pred. No. 1.3e-09;
 Matches 443; Conservative 219; Mismatches 753; Indels 715; Gaps 100;

QY 370 RVEKGLKLTITGAQTSALTIDNIGVONGDLKVQ-----LAETLT--SLKAVTTEN 422
 Db 3 RKKKGL--SILLAGYVALGLAVLPGEIOLAKAISQIKGSSLANVYAAATKQATPT 60
 QY 423 LIANEKYTVGKRLITDKIGFTNDMGIDSKPYLD----- 458
 Db 61 TAAFTNQAIA--TQAAAGIDY--NKLNVQOODIYVDVIVQMSAAPASENGILTDSSTA 118
 QY 459 -----KDTGHAGGOKITKLAVDDAATYG-----QLKRVNQT 494
 Db 119 EIDQETNKVIAQASVYAAAEVYQQTAG-----ESYGVNGFSTKYRVVDIPKQI 172
 QY 495 AESALQETVYKVY---DKNGANDSKITTYGKNNKPDQVNTFLKXGENDVYT--- 547
 Db 173 A--GVKTVITLAKYIYPPDAANSANVQAVMSYKTKGETVYSV---DSGIDPTKHD 226
 QY 548 -----TET-----NGYTFGLN--QNN----- 562
 Db 227 MRLSDKDVKLTKSDVEKFTDYVHGKRYFNSKYPGFYNDNDITITDDKVDQHGMYHA 286
 QY 563 GLIVGNTLNNDGSLVKNNTSNKOI---QVADGIFPTDISNKPAGIENTRITRD-- 617
 Db 287 GIIGANGTGDDPAKSVGVAPAEQILAMKYFSNDSITAKTGSATVVSALIDSAGADVL 346
 QY 618 GIGFANNNGSLDANKPRLTPTGIMAGKELTNVQSAINPATNGOLDENR----- 668
 Db 347 NMSIGNSNGNOTLEDPELAAYQ--NANSGTAAYIASNSGTSAGATGVKKDYGLDNE 405
 QY 669 -----LSTANTERSGSAATIKDLYNLSCVPLT-----FAG----- 698
 Db 406 MVSFGTSRGATVVASAENDVITQAVTITDGTGLQIGPEITQLSHDFGSPDKKFXI 465
 QY 699 --DGPVNTK-----KIGELTKYKGGKTTADDLTK-----NNGVAVDSNLSLT- 741
 Db 466 VKDASGNLSKALADYADAKGKTAIVKRGFSFDDOKTAQAAGAGLITVINDGTATP 525
 QY 742 --VKLAKTIS--DLDVNTKTLTASDKVYVDSGNN-----TAKLONDULTESKON--T 788
 Db 526 MTSIALTITPTTPELSSVTGOKLV--DWTIAHPDDSLGVKTLTALMLNQYTEDKMSDF 583
 QY 789 GATPATNSKITGVDELKFTDNNGIALDGTITYTDKYGFAKQDGL----- 834
 Db 584 SYGPVSN-----LSFKPDI--TAPGGINWSTONNNGYTMSGTSNAPSFIAGSQALLK 634
 QY 835 ----DKSPYLDKDKLNG-----EVEITT-----NGINAGKALTGSLTITDFTNA 878
 Db 635 QALNKNPNFYAYYKQKLTALTPLEKTVEMNTAQPIINDIN-----YANVVSRRQ 686
 QY 879 TTGHVYTOGLIVDSTDKTRFASIGDVNLNAGFNKNGN-----DAKDFVSTYDVFINGNAT 934
 Db 687 GAGLVDAVALIDALEKNSYVAE-----NGYPAVELKQFTSDKTFKLTFTNRT 736
 QY 935 TAKVTV-----DGKA-----SKVAYDVNVDTTI----- 958
 Db 737 THTLYYQMSNTDITNAVYTSATDPNSGVLYDKIDGAIKAGSNITVPAGKTAQIEFTLS 796
 QY 959 -----HLTGADGNKNOI-----GVK-----TT 975
 Db 797 LPSKSDQOOEYBGLINFGSDGSRNLNLYMGFFGDMDNGKTVDSLNGITTPYPAGNGEIV 856
 QY 976 TLTK-----TDAGK-----DKAIFSVSGDDKALIN----- 1002
 Db 857 PLTKKKNGTQYGGWYVDADGKNKYVDQALAF--SSDKNALVNDISMKYLLRNISNV 913
 QY 1003 AKDIAD--NLNLTAGETRNK-----GTADPTAL 1028
 Db 914 QVDILDGGGNKVTLLSSSTNRKTYNNAH50QYIYNAAPADGTYVDQROGNIKTADGS 973
 QY 1029 GPEOVKAYKENDG-----DNDADTIT--GKDAKTNQVNT--LKLKGR---NGL 1071
 Db 974 YTYRISGVPEGGDKQOVDPFKDKSKAPYRHAVALSKTENGTQYLLTLEAADDLSGL 1033
 QY 1072 DIO-----TNKDKIVTFGINTQSLKAGNNTTLNNGSLIKTNAGNEQIOVGADG 1121

```

Db 1034 DATKSVKTEINEYTNLDAFTAGTADGY-----TKLEPLSDROAOA----- 1077
QY 1122 VKFAKNNNGVAGIDGCTGTRIRDELGFAGTNGSLDKSPHLSKOG--INAG--KKTIN 1177
Db 1078 -----LGNG-----DNSSELYITDASNAATODASVQPGSTSPDLVINGGIDPKTSS 1126
QY 1178 10SGEIAONSNDNAVYTGKTYDKTELENKISSTAKTAQNSLHEFSVA--DEQNNFTVSNP 1236
Db 1127 TTTGYEANTQ-----GGGYTFSGTYPAAVDGTYTDAQCKKDDLWMTYTAALNSFTASMP 1181
QY 1237 YSSYDTSKSDYITPAGENGIT-----TTKNKGV-----VAVGIDQTKGLT----- 1277
Db 1182 VYNADYAAQVDYADKAHTQQLKHEDTKYRLMAPTTDLKFNNNGSDQSEATIKYGVYS 1241
QY 1278 --TPKLTGNNNGKGIIVDSONGONTIGLS--NTLANVTNDKDSVTRTECGKLIINDE 1332
Db 1242 ADTKYVNVGHYVA--ALDAQHHSVDYVANGDNITIKYATNDKQNTTTECKTTTSYD 1298
QY 1333 KTRAASTVDYLSAGFNLONGEAVDFVSTYDT--VNEFADG--NATYAKVYDTSKTSKYV 1389
Db 1299 P-----DMLKSV-----TFDQGVKFGTKKFMATSAKF--YDPKTIATIT 1337
QY 1390 YQVNVDDTITIEYKDKKLGVK-----TTTILTSTGTGANKFAL-----SNOATGDLVKAS 1438
Db 1338 GKVKHPTTLQVDGKQIPKIDDLTFSEFTLDLGLQKPGVYVGDPTQNKPFQELAFSL 1397
QY 1439 DIYAHNLTLGDIOTAKGASQANSAGYVDADGNKYVDSTDNKYKQAKNDSTVKTREV 1498
Db 1398 DAVA--PTLSLDSST-----DAPYITNDPNEFOITGAIIDNAOYLS----- 1435
QY 1499 AKDKLVAQOTPDGTLAOMNVKSVINK--EQVNDANKKOG--INEDNAFVK--LERAA 1551
Db 1436 -----LSINGSSVASQYEDININSGKPGHMAIDGPVKLEKKANLVYAV 1479
QY 1552 SONKTNAAVYGDNLNAVQOTPLTAGDGTAKKLGELITKGGOTDNKTNNIGYV 1611
Db 1480 TQSEDN--TTTKNITVYEEPKTLAAPVTPS-----TTEPQOYV--LTAN----- 1522
QY 1612 ACTDGTFLAKDLNLNLSNVNAGTKRID--KGVSPVSSGQAKANPTVLSANGLDGG 1668
Db 1523 AANTGETVOYSD-----GCKTYQDVPAAGVT--ITANGTEFKSIDLYGNESPAVD 1572
QY 1669 KYISNVGKGTQDANVQOL--NEVRNLLGLGNAGNDNAGNVNIADIKDPNSGSSN 1727
Db 1573 YVVTNI--KADDEPAOLOAKOELTNL-----IASAKTILSAGSKYDD 1611
QY 1728 RTVIVAGTVLGKGNNDDEKATG--GVQVGYDKNGNANGDLSNWVTKQDGSKKALLA 1785
Db 1612 ATT-----TALAANA--TOKAQOTALDQTNASVDSLGTANRDLOT-----ALNOLAA 1654
QY 1786 TVNAGQTYLTNNPFAEIDRINEGIRFEFHNDQNEPVPVQGRNGIDSASGKHSVAIG 1845
Db 1655 KLPAPAKKTSLSL-----NOLQSVKDALGTLGQNOTDPSSTCKT 1690
QY 1846 FOAKKADGEAANAIGROTQAGNOSIAIGNAOATGQOSIAIGTVNNAKSHGATGDPSTV 1905
Db 1691 FTAALDDIYVA-----QAQAGTQT--DDLOQAT--LAKILDEVILAKIAGS-----I 1731
QY 1906 KADNSYGVGNNGQFTDQOTDVEFGVNNITVYESNSVALGSR-----SALSAG 1933
Db 1732 KATPAEAVGNMK--DAATGKTWADLADITLISQASADASDKLAHLQALSLKTKVAAA 1788
QY 1954 THAGTQAKKSDGTAGTTAGATGTVKGFACOTAVGAHSVQ--ASGAERRIONVAAGEV 2010
Db 1789 VEAATKTVKGDGDTTSDKGGGQGTAPAPAPBDTGKDKRDESGPSGSG--NIPKTPATTT 1846
QY 2011 SATSTDAVNGSOLYKATGCIANATNEIDLR 2040
Db 1847 STTDDTTDRNGQLTSKGALPKTGETTER 1876

```

RESULT 11

```

PIP_LACLC
ID PIP_LACLC STANDARD; PRT; 1902 AA.
AC P15273.
DT 01APR-1990 (Rel. 14, Created)
DT 01APR-1990 (Rel. 14, Last sequence update)
DT 01NOV-1995 (Rel. 32, Last annotation update)
DE P1-type proteinase precursor (EC 3.4.21.-) (Wall-associated serine
DE proteinase).
CN PRP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid pMW05.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
ON NCBI_TaxID=1359;
RX 11.
RP SEQUENCE FROM N.A.
RC STRAIN=MG2.
RC MEDLINE=88149035; PubMed=3278687;
RA Kok J., Leenhouts K.J., Haandrikman A.J., Ledebor A.M., Venema G.;
RT "Nucleotide sequence of the cell wall proteinase gene of
RT Streptococcus cremoris MG2."
RL Appl. Environ. Microbiol. 54:231-238(1988).
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC INSULIN B-CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein. Cell wall.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch.)
CC -----
DR EMBL; M24767; AAA1677.1; -.
DR HSBP; P00782; 1S01.
DR MEROPS; S08.019; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR003137; PA.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 3.
DR PRINTS; PR00723; SUBTILASIN.
DR PROSITE; PS00136; SUBTILASE_Asp; 1.
DR PROSITE; PS00137; SUBTILASE_His; 1.
DR PROSITE; PS00138; SUBTILASE_Ser; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW Hydrolyase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
SO SEQUENCE 1902 AA; 199910 MW; 2901C7F19B2E5DOB CRC64;

```

Query Match

4.1%; Score 437.5; DB 1; Length 1902;

RESULT 12
 P2P_LACLC STANDARD; PRT; 1902 AA.
 AC P13293;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE PII-type proteinase precursor (EC 3.4.21.96) (Lactococpin) (Cell wall-
 DE associated serine proteinase) (LP151).
 GN PRT.
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
 OG Plasmid pLP763.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Lactococcus.
 OC NCBI_TaxID=1359;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCDO 763;
 RX MEDLINE=89313288; PubMed=2501630;
 RA Kikuchi M., Ikemura H., Shimizu-Kadota M., Hirashima A.:
 RT "Molecular characterization of a cell wall-associated proteinase gene
 RT from Streptococcus lactis NCDO763.";
 RL Mol. Microbiol. 3:359-369(1989).
 CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
 CC GROWTH OF THE BACTERIA ON MILK
 CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
 CC specificity, although some substrate preference have been noted,
 CC e.g. large hydrophobic residues in the P1 and P4 positions, and
 CC pro in the P2 position. Best known for its action on caseins,
 CC although it has been shown to hydrolyse hemoglobin and oxidized
 CC insulin B-chain.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X14130; CAA32350.1; -
 DR PIR: S06997; S06997.
 DR HSSP: P00782; 2SRT.
 DR MEROPS: S08.019; -
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR003137; PA.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR Pfam: PF02225; PA; 1.
 DR Pfam: PF00082; Peptidase_S8; 3.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
 DR Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
 KW Transmembrane.
 FT SIGNAL 1 33
 FT PROPEP 34 187
 FT CHAIN 188 1902
 FT DOMAIN 188 1876
 FT TRANSMEM 1877 1895
 FT DOMAIN 1896 1902
 FT ACT_SITE 217 217
 FT ACT_SITE 281 281
 FT ACT_SITE 620 620
 FT DOMAIN 1867 1872
 FT POTENTIAL.
 FT POTENTIAL.
 FT PII-TYPE PROTEINASE.
 FT EXTRACELLULAR (POTENTIAL).
 FT MEMBRANE ANCHOR (POTENTIAL).
 FT CYTOPLASMIC (POTENTIAL).
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 FT PROTEINS.

SQ SEQUENCE 1902 AA; 200139 MW; 4B8DB844D8CDF7 CRC64;
 Query Match 4.1%; Score 437.5; DB 1; Length 1902;
 Best Local Similarity: 21.2%; Pred. No. 2.4e-09;
 Matches 457; Conservative 201; Mismatches 728; Indels 773; Gaps 105;
 QY 370 RVEKGIKGLTITGGAGNSALTDHNIGVQVNDGDKVQ-----LAETLT-SLKWVTEN 422
 DB 3 RKKKGL-SILLAGIVAGLALVLPVGEIQAKMAISOQTKSSLANVTAFAQAQATDT 60
 QY 423 LTANEKVTGKTRLTDTKIGFTNDNGIDSKPYLD----- 458
 DB 61 TAATTNQAIA-TQLAAKGIDY-NRLNKVQODIYDVIVQSAAPASENGTLRTDYSSA 118
 QY 459 -----KDTGIHAGGQKITLTNGVYDDAATG-----QLKKNQT 494
 DB 119 ETQOETNKVIAQAASVKAABOVYTOQTAG-----ESYGVVNGFSTKVRVVDIPKLOI 172
 QY 495 AESALOTFTYKVV-----DKNGNDANDSKITVGKNNKPDGTQVNTLKLKGENGVDT-- 547
 DB 173 A-GVKYVTLAKVYPTPAKANSMAVQAVSNIKYKGEIVSVI-----DSGIDPTHKD 226
 QY 548 -----TEF-----NGVTFGLN-QNNGLTGVNSTLND-GLSVK 579
 DB 227 MRLSDDKVYKLTLSDEKFTDAKHGRFNSKVPYGFVYANNNTITDDTVDRQGHMVA 286
 QY 580 NTNSKQIQVAGD-----ITFDISNKRGA-----GIENNT 612
 DB 287 GI-----IGANGDPPAKSVGVAPAEOLAMKVPFNTSATGSAIVSAIESA 339
 QY 613 RTTRD-GIGFANNNGSIDANKPRLPTGINAGKGLNVOALNPATNGQGLDFMNR-- 668
 DB 340 KIGADVYNKSLGSDSGNOTLEDPELAAYQ-NANESGTAIVISAGSGTSGATGVNXY 398
 QY 669 -----LSTANTEKSGSAATIKDVLNLSQVPLT-----FAG-- 698
 DB 359 YGLDNEMVCTPGTSRGATVVAASANTDVITQAVTITGCTGLGPEITQLSNDPFGSF 458
 QY 699 -----DGPNTK-----KIGELIKVKGKGTITDDLT-----N 727
 DB 459 DQKRYVYKADSGNLSKRVADYADAKGIAIVKRGELTPADOKQAQAAGALITVN 518
 QY 728 NIGV-----VADSTNLSLYKLAKTL----- 748
 DB 519 NDTATPTVSNALTTTPTFGLSVTGKLYDWAHHDLSGKIALITLVPNQKTEDEK 578
 QY 749 -SDLDVAV-TKTLTASDKRVYVDSGNNTAKLQNGDLTFSKQNTGATP-ATNSKTIGVDGLK 805
 DB 579 MSDFTSYGPAVSNLSFKPDITAPGGINWSTQNNNGYTNMSTGSMASPFIAQSALLKQALN 638
 QY 806 FTDNNGIA-----LDGTYITDKVGFAPAKQDSLDKSKPYLDKDLKVEVIT--NG 857
 DB 639 NKNNPYAVYKOLKGTAL-----TDELKT--VEMNTAOPIND 673
 QY 858 INAGKATLGSLNTLTDTATNATGHTVQLGIVDSTDKTRAISIDVYNAGFNENKNG-- 914
 DB 674 IN-----YNNVYSRRGAGLVYKKAIDALEKNPSTVAA-----NGYPA 715
 QY 915 -DAKDVSTYDTPVFINGNATTAQVY-----DGA-----SKVAVVNDGTTI 958
 DB 716 VELKDFSTDKTFLFTFNRTTHELTQMDSNTDTNNAVYTSATDPNGSVLYDKKIDGAAI 775
 QY 959 -----HLTGADG----- 965
 DB 776 KAGSNITVPAGKTAQIEFTLSLPKSPDOQDFVEEGLFNKFGSGRLMLPYWGFEGDNDG 835
 QY 966 -----NKNQIGVKTTLTKTPAK-----DKAIFSVNSG 995
 DB 836 KIVDSLNGITSPAGNGRGVPLLTNNK-TGTQYGGGVNDADAGNQYVDDQATIAF---SS 891
 QY 996 DDKALIN-----ARDIAD-----NUNTLAGEIRNKK----- 1021

| Query Match | Score 435.5; DB 1; Length 1654; |
|--|---------------------------------|
| Best Local Similarity: 20.6%; Pred. No. 2,4e-09; | |
| Matches 376; Conservative 184; Mismatches 600; Indels 663; Gaps 86; | |
| 630 ANKRLPTPTGAGGELTNVQSLNPATNGQQLDFMRLSTANTEKSSAAATIKDL.YN 688 | |
| 2 AKQPNFLKLIKISAG--LWTASTATIVASFAGSA--MGAIDQONRTTNGAATTVGAGFD 56 | |
| 689 LSGVPLTFAGDIGNPNTKKLGLGELTKKGGKTTADDLTKNNIGVADSDTNSLTVKLAKTL 748 | |


```

Db 57 QTRAP-----ANGVALNMI-----IANANGINFTPAGSEGLL----- 93
Qy 749 SDDAVNTTLTASDKVT-----VDSNNNTAKLQNGDLTFSKONTGATPATNSKTTIGV 801
Db 94 --LNTANNLAVVSEDDTLGFTITVNHNAHSEFLTLNACKTLTLTGQGTNNQAATKNA 151
Qy 802 DGLKFTDNGIALDGTFTYTKKVGFAKODSL--DKSFPYDKKLLKAYEVEITTINGIN 859
Db 152 QNVVQFNNGAALDNDNDLKGVIDGAPASTLVENLAPPTQKAPLLIGDAVLANGVN 211
Qy 860 AGCAATLGSNTLTDATNATGHTVLTQIGIVDSTDKTAAISIDVYNAGFNLKNNGAKDF 919
Db 212 ----GLTAVTNGFIQVSNKSPATVKAINIADGO-----GIINTDAN----- 249
Qy 920 VSTYVDVDEINGNATTAATKATYTGKASKAVYDVNVDGTTIHLTGADGNKQIGVKTTLTK 979
Db 250 -----NANP-----LNLAGGTTINFTGDTGTC-----RLVLSK 279
Qy 980 TPAKGDKAIFSVN--SGDDKALINAKDIADNLNTLA--GEIRTKGTADTALQTPQK 1034
Db 280 HAA-----ATNFTITGSLGNLKGVT-----EFNTVAVDGOILTANAGANAVI----- 322
Qy 1035 KVKENGDDNDADDTITVGDAKTNOVNTLKLKGNGLDIOTNKGDTVFEGINTOSGLKAG 1094
Db 323 -----GTNNGAGRAAG 333
Qy 1095 NNTTLNNGSLSTKNTAGNEQIOVGADVKFAK-----VN-NGVAGAGIDGTTT 1141
Db 334 FVVSVDNGKVAAT-----IDGOVYAKDWNIOANATGQVNFRIYDVGDGTT- 380
Qy 1142 ITRDEIGFAGTNGSLDKSPHLSKDGINAGKKITINOSGETAIONSNDVATGKLYDLT 1201
Db 381 -----AFKTAASKVT-----ITDSSNGNDPFGMLAOT 409
Qy 1202 ELEKNISSAKTAONSLEHFSVADGQNNFT--VSNPYSSVDTSKTSVDTT- 1251
Db 410 KVPNAITTL-----GNFTGDAASP-----GNTAGVITFDANGTLES 445
Qy 1252 -----AGENGITTKVKKGVVRVIGIDQTKITTPKLVNANN-----GKGIYDSQNG 1299
Db 446 ASADANNAVNTNITTAIEASG--AGVVLDSGTHAELRLGNAGSLFKLADGVITNGKXNQ 502
Qy 1300 NTITGLSNTLANVNTDKGSVRTTEOGK-----IKDEKTRAASIVDVLASG-- 1346
Db 503 TALVGALAAAGITLIDGSAITITGIGNAGGAALORITLADAKKTLTLGGANITIGAGG 562
Qy 1347 -FNLOGNEADVSTYD--TVNFADGNATTAATYDDTSKTSKVVYDVNVDTTIEVKD 1403
Db 563 TIDLOANGGTIKLSTONNIYVDEPLATIDOTGVDASSLT-----NAQTLLTINGKI 615
Qy 1404 KKLGVKTTTLTSTGAKKAPLSN-QATGDALVKASDILVALHNLISGDIOAKKA--SQ 1459
Db 616 GTIGANKTKLOOFNIGSSKTYLSNCAINAVINELVGN-----GAVOPAHDTYLLTR 666
Qy 1460 ANSSAGYVDADGNKVIYDSTDNKYYQAKNDGTVDKTEKAVAKDLVQAQOTPDGTLAQMNV 1519
Db 667 TTNAAGQ-----GKIIFN-----PVVYNGT-----LAAGTINLGSATNP--LAELNF 706
Qy 1520 KSVINKBOVDANKKOGINEDNAF--VKGLEKASDNKTKAAV-----TVGDINAAVAGT 1572
Db 707 GS-----KGVNVDTVLVNVEGVNLVATNITTDDANSGFEVFNAGTINIVSGT 753
Qy 1573 P-----LTFAGDTGTAKKLG-----ETLIRKGGQDNTNKLJDNNTNGV 1611
Db 754 VGGQOGKFNVALENGTTFVFLGNATPNCNTTAAINSTLIDIGENYADCAVASADGTGIV 813
Qy 1612 --AGTGTTVKLANDLNLNSV-----NAG--GTRKID----- 1639
Db 814 EFVMTGPITTVLNRQAAPVNLKQITVSGPGNVVINEIGNGNHGAVTDTIAPENSISLG 873
Qy 1640 -----DKGVSVVDSG-----QAKANTPVLANSGLDLSGKYSVNGKGTKTDAAANVOQLN 1690

```

```

Db 874 AVFELPRCIPTPNQNTMPLTIKSTVGKNTAKGFVDVSVVLGV-----DSVIADGVIG 928
Qy 1691 EYRNILGLGNAGNNA-----DGQOVNIADIKKDPNNGSSSNRTIYAGT 1735
Db 929 DONNVYGLG-LGSDNGILVNAITLYAGISLTNNNGVITLSCGVPNTPG--TVYDLGT 983
Qy 1736 VLGR-----GN-----NDTEKLAVGOVGVYDKG-----NANGD 1766
Db 984 GIGASKFQVFTTDDYNNLGNINATNATINDGVVTTGGI-AGIGFGKLTLLSGVNGCN 1042
Qy 1767 -----LSNVWVTKOKGSKALLATYNAAGCTNNTLNNPARELIDRINNGIRFEVND 1819
Db 1043 VRFADGILSN--STSMGTITKA-----NNGVITYLGN-----AFVGNT 1078
Qy 1820 GNOE-PVVOGR-NGIDSSAS-----GKHSVAI-----GEOAKAD-- 1851
Db 1879 GDSPTPVASVRFPTSDGAGLOGNITYSQVIDFGYNNLGIYNSNITLLGGTTALNGKIDLV 1138
Qy 1852 -----GEAAVAIGRO--TQAGNOSIAIGDQAQ- 1877
Db 1139 TMTLTFASGTSTWGNNTSIETTLTLANGNIGHIVILEGAVNTTTGTITIKYODANMAN 1198
Qy 1878 -TGDSIAIGTGNVYVAKHSGAIGDP-----STVRADN--SYSGVNNQFTDA 1922
Db 1199 ESGTOTYTLLOG--GARFNGLTOSPRFAVYTGSRNRPVYSLIRANODYVITRTNAENV 1255
Qy 1923 TOTDVF-----GVGNNT--VTESNSVA--LGSNSAISA----- 1952
Db 1256 VTNDIANSFPGAGVGOQNTVTFVANATNTAAYNNLLAKKSANSANFVGLVDTSAIT 1315
Qy 1953 -----GTHAGQAKSDGTACTTTTACATGTGVKFAQOTANGAVSAGSA 1998
Db 1316 NVOLDLAKDIOAOLGNRLG--ALKYLGTPETAENAG-----PENAGISAAPAAG 1362
Qy 1999 ERRIONVAG-EVSATSTDA-----VNGSOLYKA--TQIGANATNEDLRIHONENKANAGI 2052
Db 1363 DEADIDNAVYGMKAPFTYTDHOSKKGGLAGYKAKTGVVIGDPL-----ANONLM--I 1414
Qy 2053 SSAMAMASMP--QATIPGSMVTG-----IATHNGQAVANGSLKSLSDNGM 2098
Db 1415 GMAIGITKTDIKHODYKKGKRTDVNGFSFLYGAQOLVKNFPAQGSALPISLQNVKKRSOR 1474
Qy 2099 VFKINGSADTQGHVAGVAGCFH 2121
Db 1475 YF-----FDANGMNSKOITAGHY 1492

RESULT 14
P2P_LACPA
ID P2P_LACPA STANDARD; PRT: 1902 AA.
AC 002470:
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PII-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (cell wall-
associated serine proteinase) (Lp151).
GN P2P.
OS Lactobacillus paracasei.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1597;
RN [1].
RP SEQUENCE FROM N.A.A.
RC STRAIN=NCDO 151;
RX MEDLINE=92381481; PubMed=1512565;
RA Holck A., Naes H.;
RT "Cloning, sequencing, and expression of the gene encoding the cell-
RT envelope-associated proteinase from Lactobacillus paracasei subsp.
RT paracasei NCDO 151."
RL U. Gen. Microbiol. 138:1353-1364(1992).
RN [2].
RP SEQUENCE OF 189-196.
RX MEDLINE=92226694; PubMed=1564442;

```


RA Naes H., Nissen-Meyer J.;
RT "Purification and N-terminal amino acid sequence determination of the
RT cell-wall-bound proteinase from *Lactobacillus paracasei* subsp.
RT *paracasei*."
RL J. Gen. Microbiol. 138:313-318 (1992).
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some substrate preference have been noted,
CC e.g. large hydrophobic residues in the p1 and p4 positions, and
CC p10 in the p2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyse hemoglobin and oxidized
CC insulin B-chain.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M83946; AAA25248.1; -
DR PIR: B44858; B44858.
DR HSP: P00782; I501.
DR MEROPS: S08.019; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR003137; PA.
DR InterPro: IPR002029; Peptidase_S8.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 3.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
DR Hydrolase: Serine protease; Cell wall; Zymogen; Signal;
KW Transmembrane.
KW SIGNAL.
FT PROPEP 1 33 POTENTIAL.
FT CHAIN 34 187 POTENTIAL.
FT DOMAIN 188 1902 PIT-TYPE PROTEINASE.
FT TRANSMEM 187 1876 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 187 1895 MEMBRANE ANCHOR (BY SIMILARITY).
FT ACT_SITE 1896 1902 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 1867 1872 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 1902 AA; 200253 MW; D8C9F38CE5DA582 CRC64;

Query Match 4.0%; Score 433.5; Db 1; Length 1902;
Best Local Similarity 21.9%; Pred. No. 3; e-09;
Matches 471; Conservative 231; Mismatches 815; Indels 631; Gaps 114;

QY 89 IAIIGDIAOALGOSIAIGDKIVH-----NSNNANIGA-----KASGNSIAIGGVLAS 139
DB 164 VDIPLKLAQAGKVTTLA--KVTPTDAKANSANVAQWNSKRYGEGTV---SVDT 218
QY 140 G-----HNSAIGSD-DLYLKRETVQIISLPIIRGOKALNDI---YOLADTMQRYRT 191
DB 219 GIDPTKMDRLSDDKDKLTKYDVEKFTDPAK--HGRYFSKVPYGFENVADNDITDDP 276
QY 192 HAGCHASTAVGAMSYAKGHSNAFGTRATAGETYSIAVGLITATAK--AASSIANGNAQAI 250
DB 277 VDRQHGK-----HVAIGIANGTGDDPTKSVGAPAEOLLAMKVFNTSDTSAT 325
QY 251 GFAT-----AVGGSTQVW---LNRGIALGFGSVGLKQNDQVNAANVAYAPDDNQPIDNR 304

DB 326 TGSATLVAIEDSAKIGADVLNMSIGSDSGNOTLEDEPIAAVQVAN----- 371
QY 305 KATPKNATDVFSIGNGNGNSIRKTIINAGASADDVAVNAQLEKAVLANRQITFKG 364
DB 372 -----ESGTAAYISAGNS-----GTSGSAQGVN---KYYGLQDENWY--G 408
QY 365 DSNRNVREKGIKLLTTGGAQTSALTDHNIQVONGDGLKVOLETLTSLKWTENTLT 424
DB 409 TPGISR-----GAT-TYASANTDIVSQ-----ATITIDGKDLQGPETIQLS 450
QY 425 ANEVYVTKRLTDTKIGFTNDMG--IDESKPYLDKDTGHHAGQKTKTLTAGVDDAA 483
DB 451 -----SNDFTSGPDKKFFVVKD-----ASGD-----LSKGA-----AA 479
QY 484 TYGQAKRVQTAESALQTFYKVKDKKNGND-----ANDSKITTVGKNKPPDQVNTL 536
DB 480 DY-----TADAKKGLAIYKRELNPADKQKQAQAAGAGLITV--NNDGATPLTSTI 529
QY 537 KIKGNGVDVTTENGVTYFGLNONG--LTGNSLTNNGLSYKNT-----NSNKQIQV 589
DB 530 RL-----TTFPTFLSSKTQKLVDTWTAHPDDSLGKIALTLIPNRYEDK 578
QY 590 GADGTFITDISN-----SKPGAGIENTRITBDGIGFANNITGSLDANKPRLTPGINA 642
DB 579 MSDFTSYGVPNSLSPKPDITAPGNIMS-----TONNNGYTMSTSMASP-----FIA 627
QY 643 GCKELTNVQSAINPATNGQUDPMNRISTANTE-----KSSGATITDL-VNLSQVPLTF 696
DB 628 GSOAL--LKQALNNKNNFPYADYKQLKGTALTDLKJVTENTTAQPIINDINYN----- 677
QY 697 AGDTGPNV-----TKLEELIKYKGGKTTADDTLKNIGVA-----DSTDSLTIV 742
DB 678 -----NVIVSPRQAGLVY---KAALIDALEKPSIVAVENGPVELKDTFTSTKTF 728
QY 743 KLAKT-----LSDLDVAVNKTITLADKVTYVDSGNNITAKLNG--DLTFSSKONT 788
DB 729 KLTFTNRTTHTHELYQMDNTPDINAATPNSGVLTKDKIDGAIKAGSDITVPAGKT 788
QY 789 GATPNTSKTIGVGLKFTDNNGLALDGTIYITDKVGFAPADQSDLSK---TYLD--- 842
DB 789 AQIEPTLSPKSFQDQQVE-----GELNFKFS--DGSRLNLTLYMGFEG 830
QY 843 --KDKLXGEVEITNGIN---AGGKAITGLSNLTDTATNATGTVHQLGIVDSTDKTRA 897
DB 831 DMND-----GKIYDSLNGITTPSAGN-----YGIYPLITKNNTGHQYGGWV--TDADGK 879
QY 838 ASIGVLNAGFELKKNNGDAKDFVSTYDVFINGNATTAKY--TYDGKASKVAYDVNDGT 956
DB 880 QIVDDQALN-FSSDKNALYNDISMQY---YLNRISWQYDILDGQGNKV-----T 926
QY 957 THTLTGADGNKNOIKVKTITLTKT--DAKGDKAINEFSVNSGDDKALINAKDIADNLNTLA 1014
DB 927 TLS-----SSNTQTKIYDAHSOKIYIYNAPADWG--TYDQDR----- 963
QY 1015 GEIRNTKGTADTALQTEFOVKKVENGGD-----DNDADFTT-VGKDAKTQVNT- 1062
DB 964 GNKK-----TADGSGTTRISGVPEGGDKRQYDVPFKLDSAPYRHALSAKTENGTO 1019
QY 1063 --LKLKG-----NGLDIO-----TNKDGTVYFGINTOSGLKAGNNTLLNNGLSIK 1107
DB 1020 YLLTAEKDGLDGLATSVKTAINEVTNLDAITFDAGTADGY-----TKIE 1067
QY 1108 NTAGNQDQVQADGVKFAKVNNGVYAGIDGTRITRTRREIFAGFNGSLDKSPHLSDG 1167
DB 1068 TPLSDSOQA-----LNGN-----DNSAEIYLLDNASNAITNODASVQKPGSTISFD 1112
QY 1168 --INAGG--KRTITNOSGEIAONSNDVATGKIYDLKTELEKIKSTAKTANSLHEFSV 1223
DB 1113 LTVNGGIDPKLISSTTTGEANTQ-----GGGTTFSGTTPRAAVDGYTIDAGKKHDLNT 1167
QY 1224 A-DEQGNFTVSNPYSYDTSKTSQVITFAGENGI-----TRKVNKV-----VRGID 1271

Db 1168 TYDAATNSFTASMAVTVNADYAAQVLDYADRAHQLLKHFDKVKVLPAPFTDLKFNNGSD 1227
 QY 1272 QTKGLT-----TPKLVGNNGKGLVDSN-----GQNTITGSLNLAN 1311
 Db 1228 QTSATITKVTGTVSSDTKTYVAGDTVA---ALDAQHHSVDVYVNGDNTIK-----VT 1278
 QY 1312 VTNDKSGVRRTEQOKIIRKDEKTRRAASIVDLVSAGFNLQNGEAVDFVSTYDT-VNFA-- 1368
 Db 1279 ATDEDEGNT-TTEQKITTSYSDP-----DYLK-----NAV-----TFDGVAFEGAN 1317
 QY 1369 DGNATTKAVTYDDTSKTSKVVYVYVNDVTIEVKDKKLGVK-----TTTITSTGTGANKF 1423
 Db 1318 EENATISAKF-YDPKVTGATTTGKVKHPPTTLQYDVKOISIKNDLTFSFTDLGLTGLQKPF 1376
 QY 1424 AL-----SNOATGDALVKASDIYAHNLTSGLDIOTAKASQANSAGYVDAGNKNYVD 1477
 Db 1377 GVVYGGDTTQKKTQREALITFLDVA--PTSLDSST-----DAPVYNDENFQITGT 1426
 QY 1478 STDNKKYVQAANKDGTVDKTEKAVAKDLVAQAOTPDGTLAGNKKSVINKEQVANDANK--Q 1535
 Db 1427 ATDMAQYLS-----LATNGSHVASQYADININSGKPGHMAIDQPVKLE 1470
 QY 1536 GINDNMFVAGLEKASDNTKNAAYVGLNVAQTPLETFACDGTGAKKLGELTITIG 1595
 Db 1471 GRN-----VLTVAVTDSNNTTTKKTV-----YIEPKKTLAAPTYPSP-----TTEP 1513
 QY 1596 GQTDYTNKLTNNINGVVAGTGTGTVKLAKDLTLNLSVNAAGTKLID--GKVSFVDSGGA 1652
 Db 1514 AKYTV--LTNA-----AAAGEVVOVSAD-----GKTYQDVPAAAGVT-VTANGTF 1556
 QY 1653 KANTPVISANGDLGKGVISNVGKTKDDANVQOLNEVRNLILGNGNDNADNGNOVN 1712
 Db 1557 KPFSTDLVGNESPAVDYVYVNI-----KADPPAQLQTAQAKALTNL----- 1596
 QY 1713 IADIKKDPNSSSSNNRYIAGTVLGKGNDEKLAGT--GVQGVKDKDNGNEDLSNV 1770
 Db 1597 ISAKTILASGKTDAT--TALA--TOKAOTALDQTDASVSLTGTGNRLQ-- 1647
 QY 1771 WKTOKDGSKKALLATYNAAGQNTYLTNPEAIDRINEGIRFFHVNDGNOEPVVOGN 1830
 Db 1648 -----AINQIAKLPLADKTSLS-----NQIQAQKAL 1675
 QY 1831 GIDSSASGKHSVAIGFOAKADGEAVALIGKQVAGNOSTAIGDMAQTDQSLTGTGV 1890
 Db 1676 GTDLQNGQTPSTGKTFALDLDVA-----QAQAGTQ--ADQLQA-----SLAKYLDAV 1723
 QY 1891 VAKHSGAIDPSTVADNSYSVGNNOPTDAPQTDVPEVGNNTYTESNSVALGSN-- 1947
 Db 1724 LAKLAE-----IKAAPPAEYGNK--DAATGKTWADADLTLSQASASDASDKLA 1773
 QY 1948 -----SALSAGTHAGTOAKKSDGATGTTTAGATGTGVGAGOTVGAASV--A 1995
 Db 1774 HIGALQSLKTKVAVAEAKTAGKGDITGTSDKGGGQGTTPAPAPDITKDKGDDESQPS 1833
 QY 1996 SGAEKRIQNVAGEVSAISR--DAVNSQLYKATGICINAMNEIDHR 2040
 Db 1834 SGG-----NIPTRKATTTSTDDTDRNGOHTSGKALPKTAETTER 1876

RESULT 15
 OMPB.RICITY
 AC P96989; STANDARD; PRT: 1645 AA.
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (TompB)
 DE (Tomp B) (Contains: 120 kDa surface-exposed protein (Surface protein
 antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide)
 OS OMPB OR SLF.
 GN Rickettsia typhi.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MILMINGTON;
 RX MEDLINE=94040787; PubMed=8224886;
 RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
 RT "Cloning and sequence analysis of the gene encoding the crystalline
 RT surface layer protein of Rickettsia typhi."
 RL Gene 133:129-133(1993).
 RN [2]
 RP PARTIAL SEQUENCE.
 RC STRAIN=MILMINGTON;
 RX MEDLINE=92114896; PubMed=1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RT "Mapping of monoclonal antibody binding sites on CNBr fragments of
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
 RT prowazekii."
 RL Mol. Immunol. 29:95-105(1992).
 RN [3]
 RP IDENTIFICATION OF CLEAVAGE SITE.
 RX MEDLINE=92104668; PubMed=1729180;
 RA Hackstadt T., Messer R., Cieplak W., Peacock M.G.;
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
 RT membrane protein of rickettsiae: identification of an avirulent
 RT mutant deficient in processing."
 RL Infect. Immun. 60:156-165(1992).
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 CC S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: I04661; AAB48987.1;
 CC InterPro: IPR003858; rompa_1ompb.
 DR Pfam: PF02708; rompa_1ompb.
 DR Antigen: S-layer; transmembrane; Cell wall.
 KM CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1 1353 32 KDA BETA PEPTIDE.
 FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
 FT CONFLICT 657 657 H -> N (IN REF. 2).
 FT CONFLICT 842 842 V -> I (IN REF. 2).
 FT CONFLICT 1071 1071 G -> A (IN REF. 2).
 FT CONFLICT 1306 1306 G -> S (IN REF. 2).
 SO SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;

Query Match 4.0%; Score 430; DB 1; Length 1645;
 Best Local Similarity 20.7%; Pred. No. 3.8e-09;
 Matches 420; Conservative 215; Mismatches 626; Indels 768; Gaps 101;

QY 227 LAVGLTATAKAASSIANGSNAQAIGFAATVAGSTQVNLNGIA-----LGFSGVQLQK 280
 Db 12 ISAGL-VTASATATVA-----GFGSVAMGAMVQNRRTTMAATVVDAGF----- 55
 QY 261 DNDVMAANVAVAPDDNPDNRKATPKFGKADVESIGNSNGDSIRKTIIVAGAGSAD 340
 Db 56 -DGTGAV-----MLPATNSVITANSNNATTFETPGNGLN-----SLFLDTAN 98
 QY 341 TDVAVNQLKRAVLANRQITFEKGDSDNNRREKGLTGKTLITIGAGQALSALDINIGYVN 400
 Db 99 ILAVYINENTTLGHVTN--VTKQGNFFNFTT--GAGKSLTIT----- 136

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:37:28 ; Search time 88.56 Seconds
(without alignments)
4145.159 Million cell updates/sec

Title: US-09-813-214A-9
Percent score: 10708
Sequence: 1 NMHIKVKVFNKATGTFMVA.....NCSADTQGHVCAAGAGFHF 2122

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL.19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 1246.5 | 11.6 | 2059 | 16 | Q9PD50 |
| 2 | 1185.5 | 11.1 | 2353 | 2 | P71401 |
| 3 | 1039 | 9.7 | 2712 | 16 | Q9F3X5 |
| 4 | 831 | 7.8 | 1953 | 16 | Q98HJ2 |
| 5 | 730.5 | 6.8 | 1299 | 16 | Q9F3X6 |
| 6 | 728 | 6.8 | 1190 | 16 | Q9PC04 |
| 7 | 722 | 6.7 | 3930 | 16 | Q98E20 |
| 8 | 681 | 6.4 | 1098 | 2 | Q48152 |
| 9 | 670 | 6.3 | 1107 | 2 | Q9F2D8 |
| 10 | 654.5 | 6.1 | 3705 | 2 | Q9F285 |
| 11 | 651.5 | 6.1 | 2240 | 16 | Q9ZD91 |
| 12 | 622 | 5.8 | 6713 | 16 | Q98U54 |
| 13 | 619 | 5.8 | 6713 | 16 | Q931R6 |
| 14 | 604 | 5.6 | 2065 | 2 | Q93DC7 |
| 15 | 600.5 | 5.6 | 3659 | 16 | Q98LN6 |
| 16 | 590.5 | 5.5 | 2271 | 16 | Q99QY4 |

| | | | | | | |
|----|-------|-----|------|----|--------|--------------------|
| 17 | 565 | 5.3 | 3241 | 2 | Q9AHF9 | Q9ahf9 fusobacteri |
| 18 | 554 | 5.2 | 2106 | 2 | Q9XC47 | Q9xc47 rickettsia |
| 19 | 549.5 | 5.1 | 2586 | 2 | Q9VTK8 | Q9vtk8 drosophila |
| 20 | 548 | 5.1 | 4919 | 2 | Q9ZHL0 | Q9zhl0 haemophilus |
| 21 | 546 | 5.1 | 1557 | 2 | Q9RN12 | Q9rn12 haemophilus |
| 22 | 539.5 | 5.0 | 2703 | 16 | Q9K0T0 | Q9k0t0 neisseria m |
| 23 | 536.5 | 5.0 | 4152 | 2 | Q9ZHL3 | Q9zhl3 haemophilus |
| 24 | 534.5 | 5.0 | 1291 | 16 | Q9ZK07 | Q9zk07 rhizobium m |
| 25 | 533.5 | 5.0 | 2478 | 2 | Q9LCH2 | Q9lch2 staphylococ |
| 26 | 530 | 4.9 | 3890 | 16 | Q99U53 | Q99u53 staphylococ |
| 27 | 527.5 | 4.9 | 5627 | 16 | Q9L120 | Q9l120 pseudomonas |
| 28 | 525.5 | 4.9 | 2478 | 2 | Q9RL69 | Q9rl69 staphylococ |
| 29 | 525 | 4.9 | 2147 | 2 | Q9L950 | Q9l950 pseudomonas |
| 30 | 525 | 4.9 | 2402 | 2 | Q9AER7 | Q9aer7 staphylococ |
| 31 | 519.5 | 4.9 | 2468 | 16 | Q9L2M3 | Q9l2m3 pseudomonas |
| 32 | 518 | 4.8 | 2747 | 2 | Q9L800 | Q9l800 aeromonas s |
| 33 | 517 | 4.8 | 2481 | 16 | Q990R6 | Q990r6 staphylococ |
| 34 | 517 | 4.8 | 3029 | 16 | Q95582 | Q95582 synechocyst |
| 35 | 513 | 4.8 | 2514 | 16 | Q9JY30 | Q9jy30 neisseria m |
| 36 | 508 | 4.7 | 1536 | 2 | Q48031 | Q48031 haemophilus |
| 37 | 508 | 4.7 | 2349 | 2 | P94750 | P94750 escherichia |
| 38 | 508 | 4.7 | 2383 | 16 | P76347 | P76347 salmonella |
| 39 | 497 | 4.6 | 1963 | 2 | Q9XC03 | Q9xc03 salmonella |
| 40 | 492.5 | 4.6 | 1618 | 2 | Q9KKB1 | Q9kkb1 rickettsia |
| 41 | 491.5 | 4.6 | 1477 | 2 | Q48028 | Q48028 haemophilus |
| 42 | 488.5 | 4.6 | 1619 | 2 | Q9KKB7 | Q9kkb7 rickettsia |
| 43 | 488.5 | 4.6 | 2035 | 2 | Q9XCJ4 | Q9xcj4 salmonella |
| 44 | 487 | 4.5 | 2479 | 16 | Q9A988 | Q9a988 caulobacter |
| 45 | 482.5 | 4.5 | 1604 | 2 | Q9KR99 | Q9kr99 rickettsia |

ALIGNMENTS

| RESULT | 1 | PRELIMINARY: | PRT: | 2059 | AA. |
|--------|---|--------------|------|------|-----|
| ID | Q9PD50 | | | | |
| AC | Q9PD50 | | | | |
| DT | 01-OCT-2000 (TREMBLrel. 15, Created) | | | | |
| DT | 01-OCT-2000 (TREMBLrel. 15, Last sequence update) | | | | |
| DT | 01-DEC-2001 (TREMBLrel. 19, Last annotation update) | | | | |
| DE | SURFACE PROTEIN. | | | | |
| GN | XFL529. | | | | |
| OS | Xylella fastidiosa | | | | |
| OC | Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; | | | | |
| OC | Xylella. | | | | |
| OX | NCBI_TaxID=2371; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=9A5C; | | | | |
| RX | MEDLINE=20365717; PubMed=10910347; | | | | |
| RA | Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M., | | | | |
| RA | Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S., | | | | |
| RA | Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briston M.R.S., | | | | |
| RA | Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H., | | | | |
| RA | Colanuto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa Neto C.M., | | | | |
| RA | Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H., | | | | |
| RA | Faciundani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A., | | | | |
| RA | Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R., | | | | |
| RA | Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A., | | | | |
| RA | Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P., | | | | |
| RA | Kriegler J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C., | | | | |
| RA | Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A., | | | | |
| RA | Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L., | | | | |
| RA | Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y., | | | | |
| RA | Merck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B., | | | | |
| RA | Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S., | | | | |
| RA | Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A., | | | | |
| RA | de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A., | | | | |
| RA | Petexoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B., | | | | |
| RA | Pequignolo R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M., | | | | |
| RA | de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E., | | | | |
| RA | da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr., | | | | |

| | |
|----|---|
| SA | da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A. |
| RA | de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubaho M.H. |
| RA | Valleda H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L. |
| RA | Zhao M.A., Zaitz M., Meidanis J., Setubal J.C. |
| RT | "The genome sequence of the plant pathogen <i>Xylella fastidiosa</i> ." |
| RL | Nature 406:151-159(2000). |
| RD | EMBL, AE003982; AB84338.1; -. |
| KW | Complete proteome. |
| SQ | SEQUENCE 2059 AA; 204035 MW; EA6G71B5DD24E10 CRC64; |

| | | | | |
|-----------------------|-------------------|--------------------|-------------|--------------|
| Query Match | 11.68; | Score 1246.5; | DB 16; | Length 2059; |
| Best Local Similarity | 24.38; | Pred. No. 5.8e-33; | | |
| Matches 578; | Conservative 344; | Mismatches 780; | Indels 681; | Gaps 108 |

| | | | | | | |
|----|-----|--|---|------------------------------------|------------|-----|
| QY | 62 | NSAVAGISISEDDG---- | KGGANNGDSSIAI----- | GDIAQA | 97 | |
| | | : : | : : : : | | | |
| Db | 36 | MAQVINDGM---- | DGGCQRIYDNGSAGSVERVATQCEDEPWYTGARFFPGSTAAEQ | | | |
| | | : : : | : : : : | | | |
| QY | 98 | LG-SOSIAIGDNKVIHNSNNNANIGAKASGNEISAIIG-CDVL---- | ASGHASIAIGSDLL | 151 | | |
| | | : : : : : : | : : : : : | | | |
| Db | 93 | QASIRNMLTIGSLVY---- | NSQOVGVNDVLPKKTYSIRNGSVITMNYVAGTANALIGS---- | 145 | | |
| | | : : : | : : : | | | |
| QY | 152 | YLKKEIVOOISELLPIIKGOKALNDIYOLADPTNLOKRTTHAQGHASTAGAA---- | MSYAKG | 205 | | |
| | | : : : | : : : | | | |
| Db | 146 | ----- | AQSSAADALKAASLATKASG | 164 | | |
| | | : : : | : : : | | | |
| QY | 210 | HEPNNFGRAPAEGTYSIYAVGLTATA--KAASSIYVNSMAOIGFAATAVAGSSTOVNLNRG | 266 | | | |
| | | : : : : : : | : : : : | | | |
| Db | 165 | AAAIAGAKASADGVDTYALASGATAGCASSIALGLASAVN--GAVAVGGALVATVPDG | 223 | | | |
| | | : : : | : : : | | | |
| QY | 269 | -TALGFGSOVLOKRDVNAANRAYAPPDNDOPIDNRYKATPKNGATVFSIGNSGNDISI | 327 | | | |
| | | : : : | : : : | | | |
| Db | 224 | AVALLGINSVA----- | STGKGISGYDP--KRTTTPSDASAANKSTLAAV-SIGDVSSINLK | 275 | | |
| | | : : : | : : : | | | |
| QY | 328 | RKKIINWAGSADPTDAVVAVALKEAVRIANR--OITFKGDDSSNNRVE----- | KGLCKTL | 379 | | |
| | | : : : : : : : | : : : : : | | | |
| Db | 276 | TRQISGLAAGTSMTDAVVAVALKVDDELASRGWMLTLASGANSGVAVAGSSYDLKNTDKNL | 335 | | | |
| | | : : : : : : : | : : : : : | | | |
| QY | 380 | TTT----- | GGAQTSALTDHNIIGVYONGD----- | GLNVOLAET-- | TLSLKMYTTE | 421 |
| | | : : : | : : : | | | |
| Db | 336 | TTTKAIGSNDVQFNLNKRVKXTTLAVGALLNTGIALGTDVLSITTGIALITDGPAYTAS | 395 | | | |
| | | : : : | : : : | | | |
| QY | 422 | NLTANEKY----- | TYGKTRLTDTDKIG----- | FTDMNGIDES----- | 453 | |
| | | : : : | : : : | | | |
| Db | 386 | GIDAGSKVISHVAGAVSET-- | STDVANGSOLANAVOYASQPVTFITNEGAVKRSLSQSV | 453 | | |
| | | : : : | : : : | | | |
| QY | 454 | ----- | KPYLDKPDG--IHA----- | GGQKITRLTAGVDD | 481 | |
| | | : : : | : : : | | | |
| Db | 454 | VISGESSSTAGTYSGGNLKSAYVDEAGRIHLOLADSPKFGNVYINNKGKISGVATGEETD | 513 | | | |
| | | : : : | : : : | | | |
| QY | 482 | AATYGOLKRVNOTAESALQTFETKRVYDKNGNANDSKITIVGKNKPKDGTQVMTLKIGE | 541 | | | |
| | | : : : | : : : | | | |
| Db | 514 | AVNFQOLSISITAVD-- | QGWTLPASSGNG----- | SKVASSG----- | TYDLKANP | 554 |
| | | : : : | : : : | | | |
| QY | 542 | NGVDVTTETNG----- | TYTFGLANQ---- | NGLYVGNSTLANNGLSVKNTNSNKOIOVGADGIT | 595 | |
| | | : : : | : : : | | | |
| Db | 555 | DG-NLTJISKSDSNDVFNLSKDKPKVGMQSGTIVVNNDGKV----- | GSDVALIGTTGLT | 608 | | |
| | | : : : | : : : | | | |
| QY | 596 | FTDIENSKRPAGIENTRIFTRBDGIGFANNTGSLDANKPRLPTGINAGGELANNQSAI-- | 654 | | | |
| | | : : : | : : : | | | |
| Db | 609 | ITD----- | GPAYTASGIDGSKVISHVAGGV | 635 | | |
| | | : : : | : : : | | | |
| QY | 655 | ----- | NPATNGOLDPFRMLRSLTANTEKSGSATIKDLYNLSQV----- | PLTFRAGDTGPNVTK | 706 | |
| | | : : : | : : : | | | |
| Db | 636 | SETSTDVANGSQL----- | NAVQOASQPVFTTGMGA--VKR | 670 | | |
| | | : : : | : : : | | | |
| QY | 707 | KLGELKAVGKGTATADULTKINICIVADSDTNSLTYKLAKT----- | 747 | | | |
| | | : : : | : : : | | | |
| Db | 671 | SLGGSVAVISGESSSTAGMVGSGNLKSAYVDEAGRIHLOLADSPKFGNVYINNKGKISGVTA | 730 | | | |
| | | : : : | : : : | | | |
| QY | 748 | -LSLDLDVAVNTKL----- | TASDK--VTVDSGNNTAKL----- | QNGDLFRFSKONTGA | 790 | |
| | | : : : | : : : | | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|------|------------------|------------|--------------|--------------|------------------------|-------------------------|--------------|---------------------|------------------|---------------|-------------|------------|--------------|------------|--------------|----------------|------|------|--------|----------|---|-----|-------|----|----|---|---|----|---|---|-------|------|---|---|---|---|---|-------|---|---|---|---|---|------|---|---|---|---|---|---|---|------|
| D | b | 731 | GTETDVAVNEPDLKSI | STAVDOQWTL | LAASGNSKVASG | CTVLDKNTDGLT | SKSGDSN | 790 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 791 | TPATN | SKTIGVDGLK | ETDNNGLADG | TTYTKD | -KVGFADQSDLSKPYLDKDLK | 847 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | b | 791 | DVFNLSKDFKDG | M----- | TSGTVVANN | DGVAVG | -----SDVALGTTGLT | 831 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 848 | VGE | VEITNGIN | AGKRAITG | -----LSNLT | DTAFNTGHTVLOLGIVSTDKTR | ASIG 901 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | b | 832 | JDDGPAV | ASGIDAGSK | YISHVAAGV | SETSDAVNG | -----SOLNAV | OVASQPYTFG 886 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 902 | DVLNAG | FNLKNGDKD | FVSTYD | TFDELTNG | NATTAKVYD | GKASKVAYDVNDGTTIHLT 961 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | b | 887 | ----- | NEGAVKRS | IGOSV | ----- | VISGRSTAG | -TSGSNLKSVDEAAG | -RIHLQ 930 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 962 | GADGKN | QIGVYK | TTTLT | DTAKGKAL | INFVNSD | DKALINAD | IDLNTLAGELRN | -T 1022 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | b | 931 | LADSK | --FG | ----- | NVYINN | ----- | GKISGYT | 952 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 1021 | KGTADT | -ALQTF | OVAKYK | ENGD | DDNDAD | TTTVG | KDATQNTV | LKLKNGKIDQTNKD 1079 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | b | 953 | AGTEET | DAVNSQ | LKSI | ----- | STAVDO | QWTLTAS | GANSKVASG | -G 993 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 1080 | TYTEG | INTQSGL | KAGNNT | TLNNGLS | IKNTAG | NEQIOV | GADG | -VKFAVNN | GVGAGID 1137 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | b | 994 | TV | ----- | ----- | DLKNTD | GNLTLSK | SDSN | DVFNLSK | DFKDGMTS 1030 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 1136 | GTTTR | TRDEIFA | ----- | GTNGSL | DKSKPH | LSDIN | AGGKRT | ITNOSGEI | ONSDNATV 1192 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | b | 1031 | GTTVANN | DGVAVG | SDVALG | TGTLGL | TIANG | PAVTA | SGIDAGSK | YISHVAAGV | SETSDAVN 1090 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 1193 | GKIVD | KTELE | ENKIS | TAKTAQNS | LHEF | SVAD | EOGNNFTV | SNPYSSYD | TSKTSOVITFA 1252 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | b | 1091 | GSOLNAV | OVASQPYTFG | ----- | NEGAVKRS | IGOSV | ----- | GESTAG | TYTS 1135 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 1253 | GENG | TTTVKNG | GVAVRG | IDQTKGL | TFPKL | ----- | TYGNN | NGKGI | VIDSONGNTITG | SNFLA 1310 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | b | 1136 | GCN | LKSV | VDPAAR | RIHQLA | ----- | DSKFG | SVYINN | GK | -----ISGYT | ----- 1175 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 1311 | NVTND | GSV | RTTEOG | KIKED | KTRA | SIYD | VLAS | FNLOG | NEAVDFE | STYDYNF | -A 1368 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | b | 1176 | ----- | AGTEET | DAV | ----- | NFSOLK | SI | STAVDO | QWTLTAS | GANSKVASG | CTVLDKNT 1222 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 1369 | DGN | NTLAV | YTD | DTLSK | SVYVYD | NVD | ----- | DTILE | VDKL | GYKTTTLT | STGIGA 1420 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | b | 1224 | DGN | LTKS | ----- | SGDSN | DVFNLSK | DFKDGMT | SGTTV | ANN | D | ----- | GVYK | SDVALG | TTG 1275 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 1421 | NKFAL | SNQAT | DAL | VYKAS | DI | VAHLN | TL | SGD | IQ | TAKAS | QANS | SAGYVAD | GKRYVYD | SDT 1480 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | b | 1276 | LTIANG | PAV | TASG | IDAGSK | YISHV | ----- | AA | GV | SETSD | DAVNSQ | LNAV | OVASQ 1326 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 1481 | NKYYO | AKNDG | VYDK | TR | EVAD | KLYA | QAP | PDG | L | LAOM | NK | SVYNN | EQ | ----- 1527 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | b | 1327 | PVTEFG | -NEGAVKRS | --LGS | SV | YISE | SS | TAGY | SGCN | LKSV | YAD | DAAG | TIHLQ | LADSPKF 1383 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 1528 | ----- | VNDANK | KOGI | ----- | NEND | NAF | ----- | VKLE | KASD | KTKYNA | AV | TVG | ----- | DL 1566 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | b | 1444 | KNTDGN | LTKS | GS | SDN | VYNNLS | SEDL | KE | SI | TYG | NTQ | LKD | DKD | VKSS | HWLLDS | SELVITS 1503 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 1567 | NAY | AQ | P | PL | FAG | DTG | TTAK | L | CE | TL | T | K | ----- | GGQ | DTNKL | DTNNIGVAG 1613 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | b | 1614 | TDGE | ----- | TYKL | AKD | LN | LNLS | VNA | AG | T | K | ID | D | G | V | SPV | SSGAK | NTFV | L | SANGLD | IGK 1669 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| O | y | 1670 | VISN | VG | K | GT | DD | TA | AN | VO | Q | LE | N | R | L | L | G | L | G | -NAG | D | N | D | AGN | ----- | NT | AD | L | K | -D | P | N | ----- | 1721 | | | | | | | | | | | | | | | | | | | | |
| D | b | 1562 | KIT | N | V | A | G | T | E | D | T | A | V | N | S | Q | L | K | S | V | E | A | V | D | R | G | W | T | L | T | A | S | G | A | N | S | K | V | Y | G | C | H | V | D | L | K | N | T | D | N | L | A | I | 1621 |
| O | y | 1722 | SGSS | N | N | T | V | I | ----- | K | A | G | T | V | L | G | G | K | N | N | D | T | E | K | L | A | T | G | V | O | V | G | V | D | K | D | G | A | N | ----- | D | L | S | N | V | 1770 | | | | | | | | |
| D | b | 1622 | SK | S | D | S | D | S | N | D | V | F | N | L | S | K | D | F | K | D | E | V | A | G</ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

[illegible]

```

OY 1030 --TFQV--KAYKEN-----1039
DB 1129 KTFKAGKMLKAKQSEKEDTYSLODTLTGTLTITLGTANGRNDGTIVNKDGLITTLAN 1188
OY 1040 ----GDDNDADTLTVGKD-----AKTNOV 1060
DB 1189 GAAGAGTASNGMTISVTYKQISAGNKEITNVKSLKTYKQDONTADEODKEFHAAVKNA 1248
OY 1061 NTLKLGKRGGLDIQTNKDG--TYTFGINTOSGLKAGNNTILNNG--LSIKTNAGNE 1113
DB 1249 NVEFEYKGAIVASAKTDNNGKHTYIDV--AEAKVGGLEKEDTGKIKLKADNTDGN 1305
OY 1114 QIOVADGVYKPAKVVNGVGA-----GIDGTRITRDEIFAGTNGS--LDKSKPHL 1163
DB 1306 LFTV--DARKGASVAKGEFNAVTDTATTAQGTANERGVYKSGNATATFEDK-----1358
OY 1164 SKDGINAGKKTITNIOGSEIAQNSDAVTKGIYDLKTELENKISSTARTAQNLSHEFSV 1223
DB 1359 -----KKVATV--GDVAKAINDATEFKV-----ENDSATI-----DDSP 1392
OY 1224 ADEQGNFVSNPYSSYDTSKTSVTF--AGENGITTKVKNKV--VRGIDOTKGLTTP 1279
DB 1393 TDDGAN-----DALKAGDTLTLKAGKMLKVRGKNTTFALANDSVKSAIVSD 1441
OY 1280 KTLVGNNGK-----GIVDSONGONT--ITGLSNTLANVTNDKGSVYTEOGK 1326
DB 1442 KISLGTNGKKNVITSDTKLGNRAKSKTGDANDIHLNGIASLTLDLNSGAT--TILGDN 1500
OY 1327 IKDEKTRASIVLVLSAGFLQG-----NGEAVDFVSTYDTVNFADGNATTAKYT 1378
DB 1501 GITDNEKRAASVYKLVNAGMVMVGVKPAASANNQVENIDFVATYDHFVSGDKDTSVY 1560
OY 1379 YD--DTSKTSKVYVYVNDVTIEVKDKLGVKTTTLTSTG--TGAN--KFLASQAT--1430
DB 1561 VESKDNKGRTEV-----RIGATSVIKDHNGKLFYKELKDANNNGVYV 1604
OY 1431 -----GDALVKASDIYAHNLTLTSGDIQTFKAGASQAN-----SSAGYVADCNKVI 1475
DB 1605 TETDGDENGLVTKAVYDAVNAKAGWRKTT--GANGQDDPFTVASCNTVTTADNGNT 1663
OY 1476 YDSTDNKYYQANKDG--TVDKTEKVAK-----DKLVAQA--QTPDGLTQNNVKSVIN 1524
DB 1664 AENV-----KANDGSITFYKVVYKVDGLKLDGDKIVADTVTLTVADGKVTAPN-----1711
OY 1525 KQOVNDANKKQGINEDNAVVKLEKASDNKTKNAVTVGDILNAVQTLPTFGDGTCTA 1584
DB 1712 ----NGDKK-----FVDASGLADALNKLSWTA--TAGKEGTGEVDPAANSAGQ---E 1754
OY 1585 KRLGELTITKGGOTDNTKLTNNIGVAGTDGFTVKLADLTNLNSV-----NAG-----G 1635
DB 1755 VAGDKVTFKAG-----DNLKIKQSGKDEYTSKLKEMLDITVEFPDANGGSGSES 1805
OY 1636 TKIDKGVSEFVDSGOA-----KANTPVLASGLDGVKVISVNGKGT-----DT 1681
DB 1806 TITTKDGLTIPANGAGAGANTANTISVTKDISAGNKAVTVVSGLKFKFGDHTLANG 1865
OY 1682 DAANVOQL--NEVRNLGLGNMGNND--ADGNQVNIAD-----IKDPRSGS-----1724
DB 1866 TVADFEKHYDNMVKDLJNLDERKADNNPTVADNTAATVGDRLGLNVISADKTTGEPNOE 1925
OY 1725 -----SSNRVYIKAGT-----VLGG-----KGNNDTEKLTATGG---Y 1753
DB 1926 YNAOVNNAEYVFKSGNGJNVSGKTLNGTRVITTEFLAKEGVYKSNFTYKYNADGETNLV 1985
OY 1754 QVG-----VDKDGANGLDSNVVKTQK-----DGSKALLATYNAAGOTNVL 1796
DB 1986 KYGDWYTSKEDIDPATSKPMTG--KTEKYKVENGVKVSANGSKTFTVTLNKGSG--YV 2039
OY 1797 TNNPAEALIRINEOGIRFTHVNDGNQEPVYVQNGIDSSSGHSHVAIGQAK-----1849
DB 2040 TGN--QVADAIKSGFEL-----GLADAAEAERKAPAESAKDKQLSKDKA 2081

```

```

OY 1850 ----ADGEAAVAIGROTOAGNOSIAIGDNAQATGDO-----STAIGTGVNAVKHSG 1897
DB 2082 EYVNAHDKVREFANGLTKV--SAATVESTDANGDVTTFVKTVDLELPTQY---NND 2135
OY 1898 AIDDPSTKADNSYSVGNNOFTDQTQDFYGVGNNTYTESNSVALGSNSAISACTHAG 1957
DB 2136 ANGNKIVKKNADGKW-----YELNADGTASKEVTLGVNDANGK-----2174
OY 1958 TOAKSDGTAGTTTGTAGCTVKVAGOTAVGASVG-----ASSAERRIONV 2005
DB 2175 VYKVTENGADKRYTNNADADKTKGEVSNKVSITDEKHYVRLDPNOSNGGVYIDNV 2233
OY 2006 AGEVASTSTDAVNGSOLYKATOGIAN--ATNELDRHONENKANAGISSAMAMASP 2062
DB 2234 ANEISATSTEDAINSGQLVAVAKGVINLAGOVNNLEGKVKYKGRADAGTASALASQLP 2293
OY 2063 QAVTPGRSVYTGCIATHNGOGAVALGISKISDNGOVFKTNGSADTPOGHGAAGVGFHR 2122
DB 2294 QATMPKSMVAIAGSSYOGONGIAGVSRISDNGKVIIRLSGTTNSOGKTGVAAGVGYOW 2353

RESULT 3
O9F3X5 PRELIMINARY: PRT: 2712 AA.
AC O9F3X5;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE MAFB PROTEIN (HSF).
GN MAFB OR HSF-1 OR PM0714.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=PM70;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins.";
RL Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang O., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida PM70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AJ277636; CAC14203.1; -.
DR EMBL; AE006108; AAK02798.1; -.
KW Complete proteome.
SQ SEQUENCE 2712 AA; 276154 MW; 3F5579D6F32FEA3D CRC64;

Query Match 9.7%; Score 1039; DB 16; Length 2712;
Best Local Similarity 23.6%; Pred. No. 4,3e-26;
Matches 650; Conservative 325; Mismatches 871; Indels 904; Gaps 136;

OY 11 KATGTEMAVAEYKSHSTGGSCATGQVGSVRTLSRARIALAVIVIGATLNGSAVAGTG 70
DB 221 EGTYYTGTI-KYFRVSTLDDAVADQ-DVAIGPKAKTEGVAVALGHDSTNMAIESIA 278
OY 71 ISEADGKGAGNARCKSIAGDIAQAGSQAIGDNKIYHNSNNANAGAKASGNESTI 130
DB 279 I-----GKKAIVATANKGIALGHEAR-VGSKQ-----DGDVYVYGASSYVLGPK-DESSV 328
OY 131 AIGDVLASGHASIAIGSDDLYLKKTETVOQISELLPIIRGKALNDIYQ-LADTN---LQ 186
DB 329 AIDDKAVSRGEASIAIGKNAITSNKDAKQIA-----KNNIILGTNAQAIASDNSIALG 382
OY 187 KYRRTAAGHASTAVGAMSTAKGHFSPNAPCTRACTGYSILAVGLATATAKAASIAVGSN 246
DB 383 NAATNKDSDSIAIGDSATETKAHSLAVGTTSKALAEALALGLKLAEKGTSSVAMGNT 442

```


Db 2368 PRNTVLDLNTDGNIVISKNTNADKHNVT---FGLADNINVKDSVYVPGKANGKPEGEAV 2424
 QY 1857 AIGRQTOA-GNOSIAIGDMAQATGDSIAI-----GTGNV-----VAGKHSQA-IGDPS 1903
 Db 2425 VINAEDEGANGKQGISI---VGKDGKDAVALISGKDDGVTIGLGPAGADGKNNALITIGVD 2481
 QY 1904 TYKADNSYSGVNNNOFTDATQTDVFGVGNNTITVESNSVALGSNSAISAGHTAQAK-K 1962
 Db 2482 SYKGLDGNCKDQNSKTRIVYTRPNCEBQV-ATMNDGLVFGADKTEHAKLGTIVYK 2540
 QY 1963 SDGTACTTTAGATGTVKGFAGQTAAGAVSVAS-----GABRRION-----VAAGEVSA 2012
 Db 2541 GDDKNIETEVAGDTIVR-LKQNDIVKGINVTENTLVKCGAKINNNVINDVADGEVNA 2599
 QY 2013 TSTDVANGSOLYKATQGI---ANATNELDRIHONENKANAGISSAMAMSPQAYTPGR 2069
 Db 2600 TSMQAVANGSOLKHKVQOVNNOATPAIKLGDHLNKYDKDLRAGIAGATVAFLQRPBEAGK 2659
 QY 2070 SMVTGGIATHNGGAVAVGLSKLSDNGQWVEFKINGSADTQGHV--GAAYG 2117
 Db 2660 STVSLGVSGYRSESAIVAGYARNSDNNKISIKLGGGMNSRGDVPNGSIG 2709
 RESULT 4
 ID 098HJ2 PRELIMINARY: PRT: 1953 AA.
 AC 098HJ2:
 DT 01-OCT-2001 (Tremblrel. 18, Created)
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
 DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
 DE MLL2848 PROTEIN.
 GN MLL2848.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 NC NCBL_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFP303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Matenabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti.";
 RT DNA Res. 7:333-338(2000).
 DR EMBL; AF003000; BAB49874.1; -;
 KW Complete proteome.
 SQ SEQUENCE 1953 AA; 184557 MW; 892BFA8B687B35E2 CRC64;

Query Match 7.8%; Score 831; DB 16; Length 1953;
 Best Local Similarity 22.5%; Pred. No. 1.7e-19;
 Matches 540; Conservative 251; Mismatches 828; Indels 780; Gaps 107;

QY 17 MAAVEYA-----KSHSTGGSCATGOV-----GSVRLSPFARIALAVIVIGAT 60
 Db 39 MAVLRRAHRYLGRKRGSGLGMIGTRVLAALGAILGIGATPALAQYAA-----GGGT 93
 QY 61 LNSGAVAGIGISEADG-----KGGANARGDKSIAIGDIAQ-----ALGSQSIATIGD 107
 Db 94 ANSAGSVAAGPSATNGLGIAGVSGATQAGIDSIAGTSANAGSONAVAGFQSIASQL 153
 QY 108 NKIVHNSNNNANIGAKASGNESTAIIGDVLASGHSIAIGSDLLYIKKETYQ-QISELLP 166
 Db 154 NSIYLQSRITAAAGTATA--QSAIGIGTDTASQVDIAIAGRSSVASAQSVAGLSAKAT 211
 QY 167 IIRGKALNDIVQIADTN-LQKYRPTHAGCHASTAVGAMSVYKSHSNFGTRATEGTY 225
 Db 212 GTGGAMALGGGTIASINSVALGVQASATGGGANALGITSVASGSGNSTAVGTSSSAAAGL 271

QY 226 SLAVGLTATAFAASSTIAGVSNQAQIGFAATAVGSGTQVNNLRGIALGFGSQVLOKNDVN 285
 Db 272 SPAGWSAVAGSDAVVALGKSANAGLNSTALGSTNASADFAIAL----- 317
 QY 286 AANRAVAPDQNOGLDNRKATFKNATDVFSIGNSGNSIRKILINAGSADTDVAVN 345
 Db 318 -----GNOAVSSGISVAAGSG-SQATG 339
 QY 346 VAOLKEAVRLANROITFGDSDNNRVEKGLGKLTITLIGAOTSALTDHNIQVQNGDLK 405
 Db 340 VS-----ATALGNMAA---TAANATFALIGAT--AGVDGTA-----IGKAN----- 378
 QY 406 VOLAETLSLKVNTTENTLITANEKTYVKTRLTTP-----KIGFTDMNG-----IDESKPYLD 458
 Db 379 ---ASAQDALMGTSARASSQAALAVGINAVATGKAVSISGSGTAVAGDAVSIGDPSYA 435
 QY 459 KDTGFIHAGGQKITKLITAGVDDAATYQGLKVKVQTAESALQTFYKVKDKNGDANDSK 518
 Db 436 SGTGAFITGGANNINASDGTATATA-----NOAGA----- 466
 QY 519 IITVKKNNP-----DGTQVNTLKGE-----NGVDYETETNGVTFGLNQNGLT 565
 Db 467 -VAIGNNKKAIGQGSVALGNGSTAGAAGLAGNVALGNQA-TAAASSGDVALGSGSVTAVA 524
 QY 566 VG--NSTLNDGLSVKATNSKKQIQVGADGI--TFDINSKPGAGIENTTRITRDSIGF 621
 Db 525 VGTFAVINGTTTYAFQGTPTSTVSIGAPGAERLTIVA-----AG-----RIS----- 568
 QY 632 ANNCSILDANKPRLTPGGINAGKELINVOGAIPATNGQOLDPMNLSTANFEKSGSAA 681
 Db 569 GSSIDAVANGSOLFATNOAVDAIGTLLNNIN-----GGGIKFFH-----ANSLDSSA 617
 QY 682 TIKDLVNLQVPLTFAGDTPNVTKKLCEILKVKGKTTADLTDLTKNNIGV--VADSTONS 739
 Db 618 TGTDSVAL-----GPTSTATASAI--AAGSANAGASAIIGTSVASALDAT 665
 QY 740 LTVKIATLSDLDVAVNKTTLTASDKVYVDSGNNRTAKIONGDLTFSKO--NGATPATNSKT 798
 Db 666 AMGFSLKASGQFSTAVGANANATLSTALGQNALASGVQATALQKANNQA-----SEA 720
 QY 799 IGVDLKFTDNNNGJALDQTTTYTKDKVGFAPAKODSLDKSRPYLDKLGVEYITNGI 858
 Db 721 LALGNSTAGNAGVALGSGSVTAVAG-----TPNAVINGTYARQGI 764
 QY 859 NAGKAITGLNLTDATNATTTGHTVTOGLIVSDTKPRAASI---GVNLAGEMLKNGND 915
 Db 765 NPASTVSIAGCAERTLTNLAAGRISG--STPDVANGSOLFATNOAVDAIGTIVANN-- 818
 QY 916 AKDPYSTDYVDFTINGNATITAKVYDDKASKVAVADVNDGTTHTLGTADCKNKGIC-VKT 974
 Db 819 ---ISTGGGIRKYPHANSTLA---DSSA-----TGTD--SVAIGPTST 852
 QY 975 TLTFTDAKGDKAINEFSVSGDDKALINAKDIADNLN-TLAGEIRNKTGATPALTQFQV 1033
 Db 853 ATASATAIAGSNA---NAGSANAISICTSSVALDATTAMGFSLKASGQFSTA----- 902
 QY 1034 KKVKEGDDDDADDTITVGGDAKTNQVNTLKLKRNGLDIQTNKDGTVTEGINTQSGLKA 1093
 Db 903 -----VGANANAFALSTAI--GONAL-----ASGVQATALQKQANASA 939
 QY 1094 GNNITLNNNGLSIKNTAGNE-QIOVGADGVCFKAVYNNGVAGIGDGTTRITRDEIGAGT 1152
 Db 940 SDALALIGAN-----STAGNAGDVALGSGSVTAVAV--GTPNAVINGTT-----YARQGI 986
 QY 1153 NGSLLDKSRPHLSKQIINGG--KKITNIOSEIIONSNDVATGKRIYDLTELE----- 1204
 Db 987 NPA-----STVSIGAPGAERTLTNLAAGRISGSDTDVANGSOLFATNOAVAIAGTIV 1038
 QY 1205 NKIS-----STAKTQNSLHESVADQGNFT-- 1332
 Db 1039 NNISTGCGIKFYFHNSTLADSSANGTDSVAIGPSTLA-TASALIAAGSNMNASGANSASAI 1097
 QY 1233 -VSNPYSSTDT-----SKTS-DVITTFAGENG-----ITTVKNGVYRVGID-----QT 1273

| ID | Q9F3X6 | PRELIMINARY; | PRF; | 1299 | AA. |
|-----------------------|---|---|--------------|--------------|-----|
| AC | Q9F3X6 | | | | |
| DT | 01-MAR-2001 | (TREMBLrel. 16, Created) | | | |
| DT | 01-MAR-2001 | (TREMBLrel. 16, Last sequence update) | | | |
| DT | 01-DEC-2001 | (TREMBLrel. 19, Last annotation update) | | | |
| DE | MAPA PROTEIN (HSE). | | | | |
| GN | MAPA OR HSE 2 OR PM1570. | | | | |
| OS | Pasteurella multocida. | | | | |
| OC | Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; | | | | |
| OC | Pasteurella. | | | | |
| OX | NCBI_Taxid=747; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=PM70; | | | | |
| RA | Henderson I.R., Nataro J.P., Cappello R., Stein C.; | | | | |
| RT | "Evolutionary origins of the autotransporter proteins.;" | | | | |
| RL | Submitted (APR-2000) to the EMBL/GenBank/DBJ databases. | | | | |
| RP | [2] | | | | |
| RC | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=PM70; | | | | |
| RX | MEDLINE=21145866; PubMed=11248100; | | | | |
| RA | May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.; | | | | |
| RT | "Complete genomic sequence of Pasteurella multocida Pm70.;" | | | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001). | | | | |
| DR | EMBL; AJ277635; CAC14202.1; | | | | |
| DR | EMBL; AE006194; AK03654.1; | | | | |
| KM | Complete proteome. | | | | |
| SO | SEQUENCE 1299 AA; 130963 MW; 8BCCE0EB66CDB428 CRC64; | | | | |
| Query Match | 6.8%; | Score 730.5; | DB 16; | Length 1299; | |
| Best Local Similarity | 19.5%; | Pred. No. 2.1e-16; | | | |
| Matches 437; | Conservative 244; | Mismatches 500; | Indels 1059; | Gaps | |
| QY | 1 MHRIYKVIENKRTGTETMAVAEYKASHSTGGSCATGQ--VSSVPTLSEFARIATLAVLYI | 57 | | | |
| DB | 1 MKKIYTLNNAATQSVMYVSELAKA-----GKNSASCKSLVNSVSFSFTLLIAVYVLS | 56 | | | |
| QY | 58 GATLNGSAVAG-----IGISEAD | 75 | | | |
| DB | 57 G-VVNAAEYVTGNTVGSDDKCYFNASOSVIGCDATYTKTKTDKNKPAKSVYIGGATN | 115 | | | |
| QY | 76 GKGK-----GAMARGDK--SLAIGDIALQSGSIALGDKKIYNSNNNMGAKAG----- | 126 | | | |
| DB | 116 DGETNVAIGAKSSSKAASIALGDNKAKLDNQAIALGQNATA--NSDWDISIGKQAGAEOT | 174 | | | |
| QY | 127 -----NESIAIG--GDVLASGHASIAISDDLYLKEETVQOI | 161 | | | |
| DB | 175 EYSAEGRNIALGDGALKRGKGVNNIATLGTSGADRLAGTH-NVLMKT--YVNAD--EAV | 228 | | | |
| QY | 162 SELLPITRQOKALNDIYOLADTNLQRYRIRHAGHASTAVGANSYAKGHEPSNAFGRPAT | 221 | | | |
| DB | 229 RSLALTGSGTJTKELN---AKETTDNKY---YIBASNTVALGTBALATQJLAVAALIGQAKR | 281 | | | |
| QY | 222 EGTYSILAVGITATKAKAASSIAVGSNNQALGFATFVAGSGTQVVLNGLIALGFSSQYLVQR | 281 | | | |
| DB | 282 FGNQSAVAVAGNGTASGQTLATATGSNHAATGSSSIALGIVNGMDT-----ARTLASD | 333 | | | |
| QY | 282 NDVNNANFAYAPDDNQPIDNRKKAFFKNGATCTVFISIGNSGNDSTIRKLIINVGAGSADT | 341 | | | |
| DB | 334 NPTIANGLSQAT-----KSAIINAGRNAKAGIT--NTVSYIGYNAGVT | 374 | | | |
| QY | 342 DAVNVALKEAVLANRQITFEKGDSDSNRVEKGLIGRTLTITGAGQTSALTIDHNIQVYONG | 401 | | | |
| DB | 375 KTDDEQ-----SDNNKKE----- | 389 | | | |
| QY | 402 DGLKVALAEFLTSLSKAVTENTLTANEKYVYAGKRLITTDKIGFTINDMGIDESKPYLDKOT | 461 | | | |
| DB | 390 -----PATDAVFIG-----NKAGYKSNQNRMOVS---LGRDS | 418 | | | |
| QY | 462 GHAGGOKITKLAVGVDDDAATYVQGLKKVNTQTAESALQFTFYKVKYDKNGNDANDSKLIT | 521 | | | |

RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Montefiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhami A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA de Souza B.R., Roberto G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Queiroz R.B., Roberio P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zhao M.A., Zatz M., Zelditch J.C., Zeng J.C.,
 RA "The genome sequence of the plant pathogen *Xylella fastidiosa*,"
 RL Nature 408:151-159(2000).
 DR EMBL: AE004017; AAF84783.1; -
 KW Complete proteome.
 SQ SEQUENCE 1190 AA; 118446 MW; 756741B0C8D787CC CRC64;

Query Match 6.8%; Score 728; DB 16; Length 1190;
 Best Local Similarity 23.3%; Pred. No. 2.3e-16;
 Matches 354; Conservative 191; Mismatches 516; Indels 458; Gaps 59;

QY 707 KIGELIKYKGGKTTADDLTKNNIGVADSTDSKLVKLAKTSLDLDVNTKTLTASDKVT 766
 DB 27 KIGSLIQ-HAIVTSAASSKK--GTQPRSNNAATKRSKALHDKROLHIVLITVLAAS 82
 QY 767 VDSGNNTAKLQ-NGDLT-----FSKONTGATPATNSKT---IGVDGLKFTDNNGIA 813
 DB 83 GYTGVAAGVYVNSDSTENCVEILGDSSTQSFHSASNDKCPDPTQREYSLFYRYMIV 142
 QY 814 LDGTYITRKDKYFAKODSILDKSPYLDKDLKGEVEITNGINAGKATIGSLNLT 873
 DB 143 LGGSLVYNGKLGIVDISA-----TYSMLRGLSI-ATMNG-SAGIDSLAIGSGGS 191
 QY 874 DATNATGHTVTOLOGIVDSTDKTRASIG-----DVLNAGFNKKNNDGADPFSTYDT 925
 DB 192 KTDGNTSGATVAGL-----KSTALGTASOSODASISGTGATGNTFALAGNGL 244
 QY 926 VDFINGNATTAATYDGRKSKAYADVNDGTTIHLTGADGNKNOIGVTKTLTKTDAGD 985
 DB 245 TSLANGIALGASSVYTRGVALGOSGLAATASGITGYD-----PVTKSTS--- 290
 QY 986 KAINSVNSGDDKALINADIDNLTLAGELRNKGTG-----DT-ALQTFPVKKYKENG 1040
 DB 291 -----TLSTSMRSTLGAIVSIG-NITSTSTQRLTGLAAGSDPDVAVVADLKLTL--- 340
 QY 1041 DDNDADTLTVGKDAKTNOVNTLKLKNGLDIOTKDKDTFGINTOSGLKAGN-NTTL 1099
 DB 341 -----AESVGGGNLTASGANSNVALESVDLK-NSDGNLLITKTTSDNVTFLNATL 394
 QY 1100 NNNGLSIKNTAGNEQIOWGADGVKFAKVNNGVAGAGIDGTTTRITRDEIGFAGTNSLDS 1159
 DB 395 KVDLSLTGTMTA-----MTTDDGTAVGK-----RVLTDSTGLVIAEG----- 429
 QY 1160 KPHLSLDGGINAGKKTITNQSGLAONSDAYTGKTIYDLKLELEKISTAKTAONSLH 1219
 DB 430 -PSVSSGINAAGOKIMNVTGTGTA---DTPDAVNFGL-----QAVSDTA----- 469
 QY 1220 EFSVADGEGNNFTVSNPSSYDTSKTSVITFAGENGITTKVKNKGVRVIGIDQTKGLTP 1279
 DB 470 ----- 469
 QY 1280 KLTGVNNGKGIIVDSONGONTITGLSNLTANVTNDKSGVRTTEQGIKDEDKTRAASI 1339
 DB 470 -----SKGNLLASGTSSNVAFCASVDLKN-----TDGNMLL-----TKAIGI 508
 QY 1340 VDVLSAGFNQNGEAVDPSTVDYVNFADGNATKATKRYDDTSKSKSVYVYVNDDTTI 1399
 DB 509 NDVT-----FNLTALAE-DSLTGTNTAMTGTGVTGVSNTV---LGSGLVITD----- 553
 QY 1400 EVADKDKLVGVTTLTGTGANKFALSNOATGDA---LYKASDIYVHLNLTLS-GDIOTAK 1455

DB 554 -----GPSVTSISGASGNOKITTNVAACTADTDAVNFSSOLAQVASTASKGNMLLAS 603
 QY 1456 GASAN-----SSAGYVADADKNKVIYDSTDKKYYQAKNDGTVDTKRYADKIVA---QAOT 1509
 DB 604 GANSNNVPEESVDLKNDSNLLITKTTDS-----NDVTFLNATLAKVDSLTTGTMT 657
 QY 1510 PDGTLAONNKSVYNKBOVDANK--KQGINEDNAFVKEKASDNKTKNAAVTVGDIN 1567
 DB 658 YDGVTSVNSNTLSTGLVITDGPVTSISG-----SAGNOKITTNVAACTADT 705
 QY 1568 AVAOPFLPFGADGTCTYAKLI-----GETTLTKGQOTDTNKLTDNNIGV---VAGDGFYVL 1621
 DB 706 AVNFSQLOAVSSTASKGNMLLASGANSNVAPEESVDLKNTDONITVYSKESGNDVLFNL 765
 QY 1622 AKDLTNINSVNAAGTKIDDKGVSE-----VDSGQAKANTPVLSANGLDGKVISNVGK 1676
 DB 766 SSSL-KLDDKLTVDGTVMTGVTGVTGSGVTLTSGMGLVITDGPVTSISGINSKOKITNVAA 824
 QY 1677 GTRKTDANVOQLNEVFNLLGLNAGNDNADGNQVNIADIKKDPNNGSSSNRTVIRAGTV 1736
 DB 825 GTADTDAVNSQLN----- 784
 QY 1737 LGKGNNDTEKLTAT--GGVGVYDKDGNANGDLSNVVWTKQDKGSKALLATTYNAAGQTN 1794
 DB 841 MAGSAGSVHYSTYDGTQ-----CGNYNGD----- 867
 QY 1795 YLTNNPAEADIRTEQGIREFHVNDGNOEPVYQGRNGIDSSAGSKHSVATGQAKDGEA 1854
 DB 868 -----GATGTRSIAGVGLTASABG 887
 QY 1865 AVAIGRQTA--GNOSIAGDNOATGDSIAIGTG-----NVVAGKHSK----- 1897
 DB 888 ATAVSGMAAGSKRSTALGRNAAVASADGSVALDGGADGARGAESYTKXSGLONNVTG 947
 QY 1898 AIGDPTVKADNSYSVGNNOQTDAI---QTVFG-----VGNNT-----TYESN 1940
 DB 948 VSVGDASGKGFRTVSNVADAKATDAVNLRLDRVADANKRYDNKTESLSEGTFEVKVN 1007
 QY 1941 SVALGSNSA--ISAGTNA-----GTOAKKSDGTA-GTTTAGATGTVKGFAGTAVG--- 1989
 DB 1008 SL--NNSATPIAGVDDTAIGVATASGADSIMKAKASADANV-----ATGNHS 1057
 QY 1990 -----AVSAGSAGERRIQWVAGEVSATSTDAVNSQLTKATQIATNLTNEDRHIQ 2043
 DB 1058 VADRANTVSVGSASERVTVAAG---TADTDAVNSQL---NQLITLAKQYDGVVGS 1111
 QY 2044 NENKANAGISSAMAMASKPOAVIPGRSVTGCITHTNGOGAVAYGLSLDNGWYVKIN 2103
 DB 1112 LRRTDGGVGAAMAIATNIPQAVIPGRGWTSYGVSSYRQSAIAVGVSSVSESGHWVERFS 1171
 QY 2104 GSADTQGHVGAAGAFHF 2122
 DB 1172 GSANTRSQVGIGAGVGYQW 1190

RESULT 7
 Q98E20
 ID Q98E20 PRELIMINARY; PRT: 3930 AA.
 AC Q98E20;
 DT 01-OCT-2001 (TREMblrel. 18, Created)
 DR 01-OCT-2001 (TREMblrel. 18, Last sequence update)
 DF 01-OCT-2001 (TREMblrel. 18, Last annotation update)
 DE MLL4444 PROTEIN.
 GN MLL4444.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OC NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,

RA Matanabe A., Ikesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsunoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti."
 RT DNA Res. 7:331-338(2000).
 RL EMBL: AF003004; BAB51100.1;
 KW Complete proteome.
 SQ SEQUENCE 3930 AA; 380662 MW; 9ACD1ACAI85BF712 CRC64;

Query Match 6.7%; Score 722; DB 16; Length 3930;
 Best Local Similarity 21.4%; Pred. No. 1.3e-15;
 Matches 589; Conservative 302; Mismatches 966; Indels 894; Gaps 133;

QY 16 FMAVAEAKSHSTGGSCATG-----OVSVRLSFARIALALVIGATLNGSAYA 67
 DB 1115 FSAVGGGTLTVTSGNVAAGLGPDEKGIQAGM-TIGAGVNFSSAVAVDGS--GGVST 1171
 QY 68 GIGISEADGK---GGANARDKSIAIGDIAQALGSOS-----IAGDNKLYHNS 114
 DB 1172 GIAISNTSGCNVTFGVVDLFRIGTALISLDNAASGYTFENGTKVDPTVVGSGGFLVQNS 1231
 QY 115 NNNANIGAKASGN-----ESIAIGD-----VLASGHASIAIGSDIX 152
 DB 1232 AATVAVGNLVTTLNAGADVSLTNNGTIGIGITINSNGTNGVYVGGSAIIV-SADIS 1290
 QY 153 LKKEVVOIEELPIIRGOKALDIQADPTNIOKTRRHAGCHASTAVGAMSAYAKGHS 212
 DB 1291 SSATAPETAVKVDGITGGSVTFGLITSNGT-----GTGVSVNTAGS-G-1335
 QY 213 NAFGTRATAGETSLAVGLATATA-----KAASSIAGVS--NAQAIQPAAT--AVAGSNO 262
 DB 1336 VGRGA-VTVSAGANGNGISGNAGSVTFCTGTNTVTLGSANAGCIFSGTNADVLGTTN 1394
 QY 263 VNINRG-----IALGFSOVLOKNDVNAANAVAYADNOPIDNRYKATPKN 310
 DB 1395 VTFGAGANOTGIDFSSSATAGFLTITGTGDLTSRGIDLSSTGNKTI-----TFAR 1448
 QY 311 GATDVESIGNSGNDISIRKIIIVGAG-----SADPTAVVAQLEKAVRLANROITF-- 362
 DB 1449 GSS-----ITNVGVELSSGGTATTS-----ANANFTFGDGN 1481
 QY 363 KGDSSNNRVEK-----GLGKTL-----TITGAOTSAITDHNIGVONG-- 401
 DB 1482 AGDGLSTISAAGGYTMTIGDPTLGNVNFDPFETSAHLASAVGTINWSQNGSYI 1541
 QY 402 ----DLKVOLETTLTSLKMTTENLTANE-----KTVGKTRLTDDK--IGFTNDM 447
 DB 1542 AAGTGD-----LSAAVTTISVQAQDALAGTQTFAFVGTVDLSSTPTLDSGQSITGCGNNT 1597
 QY 448 N-GIDESKPYLDKDTGIIHAGGOKITKLTAGVYDDDAATGOLKKNVOTAESALQ----TF 502
 DB 1598 SITVSCTIQPVAVOGLGASGCVITG-MECVY-TGTGDLHLHLSQVQVNTAFNFSGAG 1655
 QY 503 TVAKYDKNGNDANDSKITTVGNKNNKPDGTQVNTLKLGENGVDTTETNGTYTF--GLN 559
 DB 1656 SVFNVDSAGGFNNAGGIYI-----EGTVSVAVAGSTAFKAVAGLD 1696
 QY 560 QNNGLTVGN-----STLNNDGLSVKN-----TNSN 584
 DB 1697 SNTISTNNVNVNAGTLLDVNGGTGNTTLRGLPNSATPGTLTGCGISITANSGLLVNFT 1756
 QY 585 KOIOVGADG-----ITFTDISNKRPGAGIENTVTRITRDGIGFAN-NTGSDA-NK 632
 DB 1757 DKVTVGCAVLSLGNMGSVITFADDDITISGA-----TAFSASGGGYVNTTGTINATNA 1811
 QY 633 PRLTPTGINAGKELTNVOSAINPATNGQOLDPMNRLSTAN-----TEKSSAATTKD 685
 DB 1812 QAAALDGIITAG-----INFASSTAFASGNGIDLQNLSTASFGCTGLNTNGSGTGFNNGS 1867
 QY 686 LVNLS--QVPLTFADGTGNTVKKGLLEILKAVGKTT-----ADDLTKNNIGVAVDSTDN 738

DB 1868 AYNLSGNAVISYSGGTIASNIGTAIVSIOELTGGSVTLSCGNTDGLAGAGNIVTGTDN 1927
 QY 739 --SLVTKAKLTISDDA-----VNTK 757
 DB 1928 GTAATVTFSGNRKQIDSGATDGVSLGNPNGLAFISGGLVITTSAGGAFHASTFGCTV 1987
 QY 758 TLT-----ASDKTVYDSGNNTAKLON-----GDLTF--SKONTGATPATNSKTIGVDGLK 805
 DB 1988 TVTKDGGANNITTTTGTG-NALNIDNVTVGVGINFDSISSNCTGTGIALNNVSGAINMG 2046
 QY 806 FPDNNGI-----ALDGTITIKDKYGF-AKODGSLDKSPYLDMDKXKEVEVET- 854
 DB 2047 AYDLBETIRKGVDSVSGTLAGSTLNTSLNGLNANAATIGDLNLSAALGVSNITAGDEVDG 2106
 QY 855 --TNGINAGKAITG--LSNTLTATNATNTGHTVTOIGI-VDSITKTRAASI----- 900
 DB 2107 GSFAGTIGIDMAGTGTGTGTGDTVNNNPAGQSTIANVGVGVQFSSATNAQVFGDGA 2166
 QY 901 -----GDVLNAGFNKKNNDADKDFVSTYDTVFT--NGNATTAKVTV----- 940
 DB 2167 GPAESSIATGGQVJHAHDTLPTSGD-----YFNDFVNSGDTNLSAVRYVYTAGGTG 2221
 QY 941 DCKASK-----VAYDVNDGT--TILHGA-----DG----- 965
 DB 2222 DCSLANPGSYLGAQASTANVYVLLDKNVNGAQETIDLSGTFPNLDDGOVLLAFKSGDAV 2281
 QY 966 NKNQIGVKT-----TTLTKT--DANG-----DKAINFSVN----- 993
 DB 2282 DVSQGLVDITSGSASAFFHTTQNTPIISAPGIDTLRPLVQSNNAATVINEAFSGTGF 2341
 QY 994 SGDDKALINADIANMTNLAGEIRNTGTADTALQTVQKVKVENGGDDADATITVG- 1052
 DB 2342 TGGIENL-----YSNVGAGAVAVNATGASSFELRN-----TTITAGG 2380
 QY 1053 --KDAKTN--QVNTLKLKGNKGLDIQTKDGTVEGINTOSGLKA--GNNTTLNNGLS 1105
 DB 2381 RALDFSTTGAPANTLL-----SIOGNLTK-----SSGSALAASSGOINISATDMSIA 2428
 QY 1106 IKNITGNCIOVADGVKFAKVN-----NGVAGADGT--TRTRDEIGFAGTNGS 1155
 DB 2429 IRSPAGNRVGTGTGGGIAFNVNRFPDSGAGTVSAGTIGMNGPARIQDGLSPANTSGT 2488
 QY 1156 LDKSPHLKSDG--IN--AGGKRTTNIQSEIAGNSN--DAYNGKRIYDKTE 1202
 DB 2489 FDLGLNLANNGGCVLVNTGAGTITLNNASAGVLTNSGAFLDPLTVHMTFASVTS 2548
 QY 1203 LENKISSTAKTAONSLHEFSVADEQNNFTVSNFYSSYDT-----SKTSVDITF-- 1251
 DB 2549 VNNPANAIA--NGI-----IFDGVAGTFVYGTITSTINTGPFIDAINVTYANTGTFQF 2600
 QY 1252 --ACENGITF--KYNKGVN-----VGIDQTKLT--TRKLVGNNGNK 1289
 DB 2601 GAVGINNGTGGGGIHVASGTGLVNTGLANIDTTSVGLSOGSGTTSFTGGLIDFTTGT 2660
 QY 1290 GIV-----IDSONGQHTIF-----GLSNTLANTYNNKGSVRTTEQKIIKDEDKTR 1335
 DB 2661 GIVGCTGMYTATAGAEVTVSTGGOAISLDNIAANVTLD-----STSSCGGTNNVSLIN 2715
 QY 1336 AASIVDV-----LSAGFNLOGNGEAVDFVSTYDVNFA-----DGNATTAKVYD- 1380
 DB 2716 VSGTVGLGTGLGASGASFVNVGTAIVTYGNGISQANNAALVSVGNA--GSLTFNT 2773
 QY 1381 --DTSKTSKVVYDVND-----PTTIEVKDKKGLVKT--TTLTS-TGT 1418
 DB 2774 GALDATNGTGLOFP-NADGYTVNFENGLTTLHGAGAGIDILNGSAGTFEFGPVTLITSPGT 2832
 QY 1419 GANKF-----ALSNQATGALVKAS-----DIYA 1442
 DB 2833 AFNLVGTGASVFGTITTOGNALVALSISGHSQGYVSFFGPMASAGTGLQDNDADGAY 2892
 QY 1443 HLN--TLSGDIOAKASOANSAGYVDADGNKVIYDSTDNKYQAOKNDGYDK--T 1495

Db 2893 NFNCTTNGC---DAGVDILNNGSGSFNFSTTTTSPCTAFNVAGSATVYAGTIT 2949
 QY 1496 KEVAKDKLVAQAOTPDGTLAQMNVKSVINKRQVNDANKKQINEDNDFVKLEKASDNK 1555
 Db 2950 QGTAGQRRIALADTTGC-----AINTTAVANGLNDTACTGCI 2966
 QY 1556 TKMAA---VTVGDLNAAVQAOTPLEFAGDTGTAKKLGETLTITKGGOTDITKLDNNIGVA 1612
 Db 2987 LIDGAAGVAVTVNNAASLTJGTGRTITLIDGANNAT---GTYTFNNVAIDPVAAGATHAFVVD 3043
 QY 1613 GTDGF---TYKLAKDLNLNLSVNAAGCKIDDKGVSPVDSGQAKANTPVLSANGLDLG 1668
 Db 3044 GDQGPANNDVSATIALNNVDITNPGGNVNIQMS----- 3079
 QY 1669 KVISVNGKGTQDPAANYQOLNEVNRLLGLG---NAGNDNAGNOY-----NIADI 1716
 Db 3080 ---AGSYTFEDNAA---TTRNDGGLGIVATSNAAGGIAAGCAWTLTTSVNNAYDL 3128
 QY 1717 KDPNNGSSNRRTYKAKATVGLGKGNNDTEKLTATGVOGVYDKDG---NANGDLSNVVK 1773
 Db 3129 TSNP---GATFNFSGGLATITTSIGLN---ATGGGVNVTGTSNTVQATGQILN-WNG 3180
 QY 1774 TOKGSKALLATYNAGQTYLNNPAEALDRINEGCIREFHYNDGNQEPVVOGRNGID 1833
 Db 3181 MIGSG---ATFESSLOSSGSIT---GLDAINTLGVSGGTENG---AVTVAGTTGVG 3229
 QY 1834 S-----SASGKHSVAIFQAKAD-----GEAVVAIGRQTOA 1864
 Db 3230 SDGIEINSSAFAFNFTSTTIAATDNGIFLISGANGAVTFSTVAINTNGQAIRLDNNINA 3289
 QY 1865 GNGSIAIGDNOATGDOS---IAIGTANV-----VAGKHSAGIDPS 1903
 Db 3290 VNVN---GSGIGAINDPGNCVDINGTGANTTAASYTKTSAGNIVNVSGETACTVAFSG 3346
 QY 1904 TYKADNSISVGNNOFTDPTDTPGVGNNTVTESN---SVAGLSAISAGTHACTQA 1960
 Db 3347 SLSATGGFANG-----IDVSGNSTAVINFSNATKILTTGANTVNLTSNTGATI 3396
 QY 1961 KKSOG---TAGT---TTAGATGTGVKFA-----GOTAVGA-----VSYG 1994
 Db 3397 NFTGGGLVINTTSGTGFNATGATVYVGTGNTTITTTGALNLANITIGSNVTFOSIS 3456
 QY 1995 ASGAEERRI-----ONVAAGEVSATSTDAVNSOLYKATQC--- 2029
 Db 3457 ANGAANGVILNNTGSGGLIYTGNGNTALGNSSGGTIONTTGAI---LLTITKPSF 3512
 QY 2030 ---IANATNE-LDHRHONENKANAGISGA-----MAMASMPQAVIFGRS 2070
 Db 3513 TNINIONMANGIKGTIVTFTLANSTINNVNTAHTTDDGVTENLSAGSGSTETNLSGV 3572
 QY 2071 MYTGGIATHNGOGAVAVG---LSKLSDNQGWFKINGSLDTGSHGCAAV 2116
 Db 3573 SITGNVLNNSWQDGISINNVTAGTISLTTGNSL---TSASFAAGSAGTAL 3621
 RESULT 8
 Q48152 PRELIMINARY; PRT: 1098 AA.
 AC Q48152;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE HIA.
 GN HIA.
 OS Haemophilus influenzae;
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_Taxid=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NONTYPEABLE STRAIN 11;
 RX MEDLINE-96332658; PubMed-8730864;
 RA Barenkamp S.J., St Geme J.W. III;

RT "Identification of a second family of high molecular weight adhesion
 RT proteins expressed by non-typable Haemophilus influenzae."
 RL Mol. Microbiol. 19:1215-1223(1996).
 DR EMBL:U38617; AAC3721.1;
 SQ SEQUENCE 1098 AA; 114100 MW; D977335A89F7333D CRC64;
 Query Match 6.4%; Score 681; DB 2; Length 1098;
 Best Local Similarity 24.1%; Pred. No. 76-15;
 Matches 278; Conservative 155; Mismatches 427; Indels 292; Gaps 41;
 QY 1146 EIGFAGTNGSL---DKSRPHLSKDGINAGGKRTTNIQSEBIAONSNDAYTGKIDYDKTE 1202
 Db 64 DANFNFTNNSIADAEKQVQAYAGKGLNLNENK---ASDKLVEDNTAATVGLRKLGVN 119
 QY 1203 LENKISSTAKTAQNSLHESVADQGNFTVSNPYSYDTSKTSYVITTAGENGITTKVN 1262
 Db 120 LSSKNGTRNEKSOQVKNADVLEFGKGVQYT---STSENGKHTTFALAKDLGVATATV 176
 QY 1263 KGVVAVGIDQTKG-LTPPKLTV-GNNNGKIVIDS--ONGQNTI--TGLSNTLAN----- 1311
 Db 177 SDTLTIGGGAAGATTPPKVNVSTTDGLKFAKDAAGANGDITVHLNGISLITDLVVS 236
 QY 1312 -VYNDKSVTTEBOGKIKEDKTRASIVDVLASGPNQ-----NGEAVDFVSTY 1362
 Db 237 PATHIDGQDSTHY-----TRAASIKDVLNAGMINKGVAGSTTGQSENVDFVTHY 287
 QY 1363 DYNFADGNATTKAVYDDTSKTSKYVYVNDVDTIEVKDKLGV-KTYTLTSTGAN 1421
 Db 288 DVEFESADDETTLTVVDSKENGKREVEKIGATSVIKERDKLFTGKANKETNKVDGAN 347
 QY 1422 KFLNSNATGDALVYKASDIYAHNLNTLSGDIQTPAKGASQ-----ANSSAGYVADGN-- 1472
 Db 348 -ATEDEDECKGLVTAQVDVAVNKGKGRKTTDANGQNDPRTVASGNTVPASNGTT 405
 QY 1473 KVIYDSTNKYYQAKMDGYDKTEYAKDKLVAQAOTPDGTLAQMNVKSVINKEQYNDAN 1532
 Db 406 AIVTNGTDG--ITVKYDAKVGDKLIDGKRIA---DTTALTVN----- 444
 QY 1533 KKGGINEDNAFVYGLEKASADNK---TKNAAVYVGLNNAVQAOTPLEFAGDTG----- 1582
 Db 445 -DGKANNPKRGVADVASDEKRLVYAKGLVTA--LNSLSWTTTAAEDGTLDNASE 500
 QY 1583 TARKLGEFTLTKGGQDPTNKLTDNNIGYVAGTGFYVYKLAKDLTNNVSV-----NAGGT 1636
 Db 501 QEVYAGDKVTFFKAGK-----NLVKQEGANFTYSLODALTLGLTSLITLGTGNNGART 551
 QY 1637 KIDDKGVSPVDSGQA--KANTPVLSANGLDLGKAYISNVGKGTQDPAANYQOL----- 1689
 Db 552 EIMKDGTLTPPACAGANNANNTISYTKKDGISAGQSVKNVYSGLKKFEGANFPLTSAD 611
 QY 1690 ---NEVNRLLGLGNAGNDN---ADGNOVNIA-----IKDPNNGSSSS--- 1726
 Db 612 NLTKONDAAKGLTNDKGTQKQTPVAVDNTAATAGVGLGGLVSAIKTGTGSTEYHD 671
 QY 1727 ---NRVYIKAGTVLGGGNNDEKLTATGVOGVVDNAGNNDL--SNVWVKTKQKSGSK 1781
 Db 672 QVRNANEVK---FKSGGLNVSGKTVNGRREITFEL---AKGEVYVSNETVYKELNGKET 725
 QY 1782 ALLAT---YNAAGQTNLTNNPAEALDRINEGCIREFHYNDGNQEPVVOGRNGIDSSASG 1838
 Db 726 SLVKGQKYYTSKIDIDLTTGQF-----KLKDN----- 753
 QY 1839 KHSVAIFQAKKADGEAVALIGRQTOAGNOSIAIGNAQAOTGDOSIAIGGNVYVAGHSGA 1898
 Db 754 ---TVAAKYQDK-----GKRVSVTDNTEATITNK---GSGYVYVGNQVADA 793
 QY 1899 IGDPSYTKADNSYVGNNO---FTDQOTDVGGVGNNTVTESSNVAGLSNSAISA 1952
 Db 794 I-----AKSGFELGLADEADAKKRAFDDTKTALSAGTTEIYNAHDKVRFANGLNTKYSA 846
 QY 1953 GTHAGTQAK-----KSDGTAG 1968

Db 847 ATVESTDANGKVTTEVKTVDVELPLTOIYNTDANGKKTTRKVKQOTRWKVELNADGTAD 906
 QY 1969 TTTT-----AGATGVKGFACOTAVGAHSV-----1994
 Db 907 MKREYTLGVNDSDGKVVVDNDGKMYHAKADGTADTKGEVSDKYSTDEKHYVSLDPND 966
 QY 1995 -ASGERRIONVAAGEVSATSTDAVNSQLYKATOGIAN--ATNELDRIHONENKANA 2050
 Db 967 QSKGKGVVDVNDVANGDISATSTDAINGSQLYAVAKGVTLNAGVNNLEKGVNKKRADA 1026
 QY 2051 GISSAMAMASPOATIPGSMVTGATIHNGCAVAVGSLSDNOQWPKINGSDTQG 2110
 Db 1027 GRASALMAAOLPQATMPGKSMVAIAGSSYQGGALGVSRISDNKVIITRLSGTTNSG 1086
 QY 2111 HGAAVAGGFHF 2122
 Db 1087 KTCVAGVGYQW 1098

RESULT 9
 Q9F2D8 PRELIMINARY; PRT; 1107 AA.
 AC Q9F2D8: 01-MAR-2001 (Tremblrel, 16, Created)
 DT 01-MAR-2001 (Tremblrel, 16, last sequence update)
 DT 01-DEC-2001 (Tremblrel, 19, last annotation update)
 DE SAPB PROTEIN.
 GN SAPB.
 OS Salmonella typh.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
 RT "Evolutionary origins of the autotransporter proteins."
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ277623; CAC14217.1;
 SO SEQUENCE 1107 AA; 113223 MW; F5C7CA651FED51AB CRC64;

Query Match 6.3%; Score 670; DB 2; Length 1107;
 Best Local Similarity 23.7%; Pred. No. 1.6e-14;
 Matches 304; Conservative 193; Mismatches 413; Indels 372; Gaps 59;

QY 993 NSGDDKALI--NAKDIADN--LNTLAGELRNKGTG---DTALQTFQVKKVKGNGDDN 1044
 Db 46 NAGNDTGSINONTDTLATNTTSTINLSNSVTTLTDALLDASGTFSSAS--NGSASK 102
 QY 1045 DADDTI--TVGKDAKTQVN-----TLKLGKNGLDIQTNRKDGTVFPGINTOSGLKAGNN 1096
 Db 103 ITMLAGTILADS--TDAVNSQLEFDTNEKYDQNTADITTN-----TNSINONTDTLATNT 156
 QY 1097 TTLNNGLSIKTNAENQIOVGADYKFAKVNNGVAGIDGTTTRITRDEIGFAGTNGSL 1156
 Db 157 TSTINN-----LSNSV-----TTTLDALLMMAAGTGF 183
 QY 1157 DKSKPHLSKGINAGGKKTINIOSGEIAONSNDAYTGKTYDLKTELENNISSAKTAON 1216
 Db 184 SASR-----NGSASKITNLAAGTLAADSTDAVNSQLEFDTNEKYDQNTADITTN 234
 QY 1217 SLHEFSVADEQNNFTVSNPYSSTYDTSKTSQDVI-----TF-AGENGITTKV--NKGVR 1267
 Db 235 SINONT--DIATNTSINNLSNSVTTLTDALLMDADSGTFSSASRSGASKITNLAAGT 292
 QY 1268 VGIDQTKGLTTPKLVGNNGKGIIVDSQ--GQNT--ITGLSNTLANVTNDKGSVFTTEQ 1324
 Db 293 LAADSTDAV-----NGSQLYETNOKYDQNTSAIADINISITNLSQD-----333
 QY 1325 GKTIKEDKTRASIVYVLSAGFVLQNGEAVDEVSTYDVFADGNATTAKYVYDDTSK 1384
 Db 334 -----NLSWN-----ETTNSFSASHGSSSTTKITNVAAGE 363

QY 1385 TSKRYVDVNDVDTTIEVKDKKLGKTTTLTSTGTANKRFALSNQATGDLVKAADIVAH 1444
 Db 364 LSESTDA--VNSQLEFETNEKYDQNTDIAANTTN-----ITQNSTAIENL 408
 QY 1445 NTLSDGIQTKKASQANSSAGYVDADGNKYIYDSTDNKYQOAKNDGTVDKTRVADKLV 1504
 Db 409 NTSVSDINT-----STGLTD--NALLMDE-DTGAFAHNGGSTSKITNVA-----451
 QY 1505 AOAOTPDGTIAQNNVKSIVKREOVNDANKKQGINEDNAFVKGLEKASDNKRTNAAVTG 1564
 Db 452 AGALSEDST-----DAVNSQLEYETNCK-----VDQNTS-----AIA 483
 QY 1565 DLNAVAOTPLTFAGDGTGTTAKLGETLITTKGGQDTNKLTDNNIGVYAGTGTGVLAKD 1624
 Db 484 DIN-----TSTNLTGTDALSDWDEGAFSASHGTSGETKITTN--VAAG-----ETASD 529
 QY 1625 LTNINSVNAAGTKIDKGV--GFVDSGGAKANT--PVLSANGLDGGKVISVNGGK 1679
 Db 530 ST--DAIN--GSQLEYETNMLISQINESISQLAGDSETTYTENGG--TGKAYIRTNNGLE 583
 QY 1680 DTDA-----ANYQOLNEVENLGLGNAGNDNADGNQVNIADIKKDPNSGSSNRVYKA 1733
 Db 584 GQDAYATGNATATVGYDAVASGACCLALGONSSSTIEGSTA-----LGSGSTNRATITG 638
 QY 1734 -----GTVLGKGNNDTEKLTATGGVOYVDKDGANAGDLSNVWTKQKQSKALL 1784
 Db 639 IRETSATSDGVLI--GYNTDRELLCALSLGT--DDESYROIITN--ADGSE-----684
 QY 1785 ATYNAAGQTYLNNPFAELDRINEGIRFEHWDNGQEPYVVGGRNIDSSAGKHSVAI 1844
 Db 685 -----AQDAVTVROIALNAGAVTTPTPKYHANSTEE-----DLSLAGTDSLM 728
 QY 1845 GFOAKADGEAAVAIGROTA---GNOSIAIGDNAQATDOSIAIGTGNVYVAGKHSAGIGD 1901
 Db 729 GAKTIYAADIGIGLNTLYMADALINGAIGSNARAHANSINGNSQTT--RGAKTD 785
 QY 1902 PSTYKAD-----NSYVGNNN--QFTD--ATQTDVFGV-----NNTVT 1937
 Db 786 YTAVMNPTQNSVGEFVSQEDGQROIITNVAAGSADYDANVGLKTAQVSNRNTOSIT 845
 QY 1938 ESN-----SVALGSNSAISAGT--HAGQAKKSDGTA--GTTTAGATGVKGA 1983
 Db 846 NLNTQVSNLDTVTNINENGIDYITGSTYKFTKNTDGDADNAGDASVLAIGSSIAAE 905
 QY 1984 GQPAVG-----AVSAGSAGERRIONVA-----2007
 Db 906 NSVALGINSVADEANIVSVSSSTQQRITNVAAGVNNTDVANNQOLKASEAGSVRYETNA 965
 QY 2008 -----GEVSA--TSTDVANSQLYKATOGIANATN-----ELDHR 2040
 Db 966 DGSVNSVLMNGDSSGCTTRIGNVAANDTDVANNQOLKRSVEANTYTDOKKGENNSK 1025
 QY 2041 IHONENKANNISSAMAMASPOATIPGSRVMTGATIHNGOGAVAVGSLSDNOQWPKING 2100
 Db 1026 IKGIENKMSGISAMAMAGIPQAYAPGANTSTAGTFNGESVVALGVSVSSGGWV 1085
 QY 2101 KINGSADTQGVHGAAGGFHF 2122
 Db 1086 KLGQTSNSQDYSALAGGFQW 1107

RESULT 10
 Q9F285 PRELIMINARY; PRT; 3705 AA.
 AC Q9F285: 01-MAR-2001 (Tremblrel, 16, Created)
 DT 01-MAR-2001 (Tremblrel, 16, last sequence update)
 DT 01-DEC-2001 (Tremblrel, 19, last annotation update)
 DE YAPB PROTEIN.
 GN YAPB.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Db 2068 VGNVTAKVIDADATLYVD-----TALLDHLVLENGTLLNVAATLTATPAEDKSGTGVXXPSGI 2123
 QY 1569 VACPTLFEAGDTGTAKLGEITLIKGGDTNKLITDNNIGVAGCTPGTIVKLAKDLNLT 1628
 Db 2124 VNLNTTFALSADNAALATL-----KLSDDSTVYGTIDRIIHL-----DL 2168
 QY 1629 NSVNAAGTKIDDKGVSPVSSGAKANTPVLSANGDILGGKVISNMGKTKPTD---AA 1684
 Db 2169 N-----GSTLL-----FDGSPPOGAN-GVYVYTDIALNSGTLISITGAGWNEHEHPYTP 2217
 QY 1685 NVOOLNEVRN--LLGLGNAQNDNADNOVNIADIKDPNSGSSSNRYIYAG----- 1734
 Db 2218 NVSLLEDRDIDILLQILDADNVTGNANDELIMINGTISAGGVOSTVQGGYTVANATH 2277
 QY 1735 ----TVLGKGK-----NNDTEKLTG-----GVQVGVKDCNANGDLSNV----- 1770
 Db 2278 NYGNTSMGSGGLVYNTLSALELLADGANMLLTBSGLTANRELNAELSGVGLVYDQ 2337
 QY 1771 -WVKTKDGSKK-----ALATYNAAGQNTYLT-----NNPAEADIRINE 1809
 Db 2338 NGALTLTANGNNRREGTTVTAGELLIGANGAFQOTSLNLTASGANININGYRQTGAVTN 2397
 QY 1810 QG-----IRPHVNDGNOEPVVGCRNKIDSASGKSHVAIGFOAKD 1851
 Db 2398 TGTVTLGNGELSTDTLINTGMINTVDG---ILNLENGASSISG-GLTGNGILINIKG 2452
 QY 1852 GEAVALIGROTOAGNOSIAIGDNAOAT--GPDSIAIGTGVANAGKHSAGIDPSTVAKAD 1910
 Db 2453 GDFLISIDNNGLAGQTIIS--DGASVTLGNGGTTIGTG--LGSSVIDYLDGLMLV--ADNS 2508
 QY 1911 YSVGNNOFTDATTQDVEFGVGNNTVYESNVALGNSAISAGTHAGTOAKKSDGTAGTT 1970
 Db 2509 LA-----NVISDG--TINTTATVTLGNSSPSGAHQIGTNGELIVGQASNL 2553
 QY 1971 TTAGATGTGAGQTAAGVAVSVA-----SGAERRIONVAAGEVSATSTPAVNSQL 2023
 Db 2554 GASSAT-----VNLGTLSHLILNGVSESIANVLG--VAGSTVLIIGAD- 2597
 QY 2024 YKATGIATANTNELHRIHONENKA-----NAGISSAMAMASMPQAYIGRSMYTGIA 2077
 Db 2598 ---TALTAANNSGFLGAYALAGNSKLTIVASTNNIGASSVALAG-----TGDLSLSGFN 2648
 QY 2078 THNGGAVAVGLSKLSDNGQ 2097
 Db 2649 GTFGNSVTGSGVLQVTDAAE 2668
 RESULT 11
 Q92D91 PRELIMINARY; PRT; 2340 AA.
 AC Q92D91;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE CDL SURFACE ANTIGEN (SCA3).
 GN R451.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_taxid=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MADRID E;
 RX MEDLINE=99039489; PubMed=9823893;
 RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
 RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Neeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria."
 RL Nature 396:133-140(1998).
 DR EMBL: AJ235271; CAI:14908.1; -
 DR InterPro: IPR003858; TomPA: TomPB.

DR Pfam: PF02708; rompa_rompB; 1.
 KW Complete proteome.
 SQ SEQUENCE 2340 AA; 247862 MW; CCA070F93C165F26 CRC64;
 Query Match 6.1%; Score 651.5; DB 16; Length 2340;
 Best Local Similarity 21.2%; Pred. No. 1.4e-13;
 Matches 519; Conservative 302; Mismatches 879; Indels 743; Gaps 121;
 Db 10 NKAATGEMVAE-----YAKSHSTG-----GSSCA-----TGQVGSVRLSPARIALAVL 55
 Db 14 SKILKFLATLACGLTLPNSNATGTIIIPNNGSVLNTDAGLVGCV-----FNNGDITOI- 58
 QY 56 VIGATLNSAYAGIGISEADGGKGANARGDKSIAIGDIAQALGSOIAI----- 105
 Db 59 ----VNGREIKISADKANALIGGINLKE-----LPFGEGVEQVNSIGPLNAGEDLN 109
 QY 106 --GDKKIYHNS-----NNANIGAK-----ASGNESI 130
 Db 110 TNEGPLKPTSNVNTSLITGVGTPTESNIDFAGKANTLQINDLNTTIXIDNTVAGNNGSI 169
 QY 131 AIGDVLASGH-----ASIAISDDLYLKKEIYQOISELPIIRGKALNDIYQLADT 183
 Db 170 TREGSGIISNHLGYNSLGLIVNGEAKIYAPEANNIT-----I 209
 QY 184 NLOKVRRTHAQHAFTAVGAMSYAGHFSNAPFRTAEGYSLAVGLTARAKASSIAY 243
 Db 210 NKNKINLTNNSILITLCOGNITTLKGNINNT--TEIDGGILINLAYDL-----GSSSIT 262
 QY 244 GSNQAQIGFAATAVAGSGTQVNIHRCIALGFGSQVYLQKNDVNAANVAYADDPNDPINDR 303
 Db 263 GD-----IG-----NISLPTINVLGSAT--FNSTILKATINIKHNTSTLNDN----- 307
 QY 304 YATFKNGATDVFSIGNSGNSDIRKIIN--VGASADTDAVNVAYOLKEAVRLANRQIT 361
 Db 308 -----IIVIGNIKGN--KDLINRVHGTNLDNEKIIIPAPOK-----THGLIN 349
 QY 362 FKGDDSN-----RVKGLGKTLITTGCAO-----TSA 389
 Db 350 FGNMTLNGNINNLNLIKFSGGHGTMLNQGNTKYDNLVFADSVLDSGTISVNGLLDTC 409
 QY 390 LUDHNIQVQNDGKLVQLAETL-----ISLKWTEENTLANKYVGTRLTLTKIG 442
 Db 410 VFNNSNV--NGSTLIINAKNTISAKLINATKAKIQIANLTMN-----PSAG 456
 QY 443 FNDMNGIDESKPYLDKDTGHAAGOKITKLAVGVDDAATYGGOLKYVNOTAASLOT 502
 Db 457 DISDIRIADNTIYTIIDAKN-----GVNLLNNNAKIIIFEGADS--MLALIN--TGYADRTF 509
 QY 503 TV-KKVDKNGDAND-----SKIITVCKNNKPDGT--QVNTLKL----- 539
 Db 510 TIYNNLNGSGNDEYGIKLEAIKIVITLANOGSPYTIQDNTHRKELIVGAGDIID 569
 QY 540 -----GENGVDTTEGTNGVTVGLQNNNGLYGN 568
 Db 570 TIFTKLINSITGQITFNKTLDLGAGNIAPFGKHTLVVNGVTSITSENNOGILTNS 629
 QY 569 STLNDGSLVKRTNSNKOVOGADGIT-----FTIISNKPQA-----GI----- 608
 Db 630 G--NITVGITHELGLKLVNIGADPYTCSANVFASVALINPSSVILLADGYTLGVEYTH 687
 QY 609 ENTTRITRDIGFANNSTGLANKPRLPTGINAGKELIVQASAINPATNGOLDFENR 668
 Db 688 NNTKGLSLGTC-SNITGIGTNSAALKINIGACA--SNIDSNIYAG----- 732
 QY 669 ISTANTEKSGSAATIKDLNYSQVPLTFAAGTGPVNTKKLGLKVGKGTTA--DOLTKN 727
 Db 733 --STVLTLDQSELTLLNDVNVNNSIITFGNGGICIGNICANGAALOEVEFN 791
 QY 728 ---NIGYVADSTDNSLTYKLAK-----TSLDLAVNNTKTLTASDKVN--VDSGNNTAKL 776
 Db 792 GTTNTIGTANSONFTVAHSAANVITGLTIGALKYKDTGTTIIGHGGLVGDIDPNNKAGKF 851

QY 777 QNGD-----LTF-----SKONTGATPATNSKTI----- 800
 DB 852 ILGGAGMIDSVLCNGVAGTLDIFIGDGNVTQINIGADNMSISTINIGDNTKNVTIAND 911
 QY 801 --VDLKFEDNNGIALDGTYYITRKQVFAKQDQSDKSPYLDKDKLQVEVEITNGI 858
 DB 912 IFVDNHFITNGILQLOGN--LFTHNIDFGANGTLEFN-----GNNVTYLNAL 958
 QY 859 NAGKATIGLSNLTLD--AINATTG-----HYVQLG-----VDSKTRAAISGDVL 904
 DB 959 IVMQO--NGILMFTNLKASDDTITGVKILINIGQIGPQNETIQQNNKLTLLVSVSSSI 1016
 QY 905 NAGFNELKNNDKADK---FVSTYDVTDFING--NATTAQVYDQKASVAVDVTGTH 959
 DB 1017 NF-----GDANSQILSPVDTQIKFNNKLNLETGGI-----ITLDSN--GNNLT 1059
 QY 960 LFGADGNKNOIGVKTTLTLPKTDKGRKAI-----NFSVNSG--DDKALINAKDI 1006
 DB 1060 ISGNNGIK--LGSKNELSLSLNIKQVTVTNDLQINIHQILNINNGALFDDQSLTSAKIK 1117
 QY 1007 ADNLTLAGELRNTKGTADALOTFOVKKKKENGDDNDA-----DTIVGDAKTNOY 1060
 DB 1118 NINGTAVAGG-----ATYTLDAINDFNLFTSGMVEKHQDSILELKNSSNTND 1165
 QY 1061 NTEKLGKNGELDIQNTKGTFTGINTQSLKAGNNTLNNGSLIKNTAG--NEOLYOG 1118
 DB 1166 HTITL--TSALDPGNNOFGIILKLTDP-----NKLTIDNG--NVAATLGTANHLK-- 1213
 QY 1119 ADGVKFAKVNNGV--VGAGIDGTRITRD--EIGFAGTNGSLDKSPHLKSDGINA-- 1170
 DB 1214 --QTFEASINDGALATKVGINVENVTFLNKIDIEVNANVAFNNKNTYATGATNINCHVD 1271
 QY 1171 --GGKRTNQ-----SGELAKONSNDAYT-----GKLYDKTELENKISTARTAO 1216
 DB 1272 FQAGAGVINLNDLIEIDGVTSTGN--VNGTLEFNSSGKVTGLINIMVLQAGADVSLS 1329
 QY 1217 SLHEFSVADDEG--NNFT--VSNPSSYDPSKT-----SD 1247
 DB 1330 ASGWSTTEILOGNKNMLTFRANSHTLTDINKTGGDLNLFVINGSSVSGSICANAAVGD 1389
 QY 1248 VTFEAG-----ENGITTKVKKGVVRGID--QTRGLTTPKL----- 1281
 DB 1390 IINAGSVNFSNFKSGNIVISDQATQVNNVNTATDISKNANNGTLKLNHTPINITS 1449
 QY 1282 TVGNNNKGIYDSONQNTT-----GLSNTLANVTNDKGSVR 1320
 DB 1450 TILGNNAIG--TIEVANDVTITGLAQONIHFSNATQATLTLGAASQVYNTTACNNIH 1508
 QY 1321 TTEOGKTIKDEKTRASIYDVLS--AGFNLQNG--EAVDFVSRYDVTVPADGNA-- 1372
 DB 1509 TLE-----VTDFDIGNDGIIDANNRLKSIELTNGYVITNSPHVYSITTANNAQGNVKL 1564
 QY 1373 --TAKVYTDTSKTSKYVYDVNDTDTI--EYDKKLGVYTTTLTSTGTANKFALSNOA 1429
 DB 1565 NIEGITYDLSKIKSLANVOISEDTTIRDDVYSKYLYINDAGKTINEDRDNNMNPRLND 1624
 QY 1430 TGDALV-----KASDVAHLNLTSGD--IOTAKASQAN--SSAGYVAD--GKNYIYDSTD 1480
 DB 1625 IPDALIDLVLPRSLSENFETDTRKADNLFPADDTATANKDVAVIDAHDINGILKEND 1684
 QY 1481 NKY--YQAKNDGTVDKTRKAKDLVAQAOTPDGTLAOMNVKS-----VINKQ 1527
 DB 1685 NAWLTQEKNNATIT--ELASDKFM-----LLQKIKAKATLILADANLVLNDVE 1731
 QY 1528 VDNANKKOGIENEDNAFVKGLEKASADKTRNAAVTVGLDVAVQPTLTFAGDTGTAKKL 1587
 DB 1732 VNTMLNVDIYDLDA--NVELKYTGWVTHNGLLTI-----ITY--FDTLAQKG 1775
 QY 1588 GETLITGGGOTDKLNDNNIGVAVAGTGTFTVAKLADLTNLS-----VNAAGTK 1637
 DB 1776 GHILVSGSNVMSDL--DNLT-----IKRASHDITNTSDTKHQIYKLETGAIYTP 1826
 QY 1638 IDKGVSFVSSGQAKANTPVLASNGL--DLGKGVISVNG--KGTKDDPAANVQ 1688

DB 1827 VPQKYLIDASEQNKFEVWVADANGVLVLLPDTGR--DDTGRRDTRGRGNTD----- 1878
 QY 1689 LNEVRNLGLGN-----AGDNADGNQVNIADIKR--DPN-----SGSSS 1726
 DB 1879 --NCCRDNCVGNISNNSSNEAGSSSDKN--YGITVVPPIFDPSPILDTKNNYASGIAN 1936
 QY 1727 --NRTVAGTVLOGKGNNDTEKLATGQVGVKDG--NANG-----DLSNWKV-- 1773
 DB 1937 QLINHVKFGNTIDGKLLNDIGFMSPNRYVTTLRLSKRINVLNGLNCSYGLNIEVEN 1996
 QY 1774 -----TOKD--GSKKALLA--TYNAGQNTVLTNNPAAIDRINDOGRIFPHV 1817
 DB 1997 FLTLDAINDNFTAKIGRLEELSDANTVNGLNTNTLNNK-----INLRL 2045
 QY 1818 NQGNDEPVYQNRGIDSSAGKHSVAIGFAQADGPAVAIGRQTOAGNOSTAIDNAQA 1877
 DB 2046 NNNQALIAAGE--DNIVTGTMGCSFYGKIKONSKNS--ASGYQNTGGILGPDYNT-- 2100
 QY 1878 TGDOSTAIGTGNVAVG--KHSGALGDPSTVAKADNSYSVG-----NNNOFTDA----- 1922
 DB 2101 --DNSTVIGAATYTMADSKYKHKNKNGDRTKAKSNITSYGLNMLTNFVEALGVYGR 2158
 QY 1923 -----TQTDYFGVGNITVYESNSVALGSNSAISACT 1954
 DB 2159 NKIKNEKRITITLTDQIAIGKFINFYSYELLGYNILSHRT 2201

RESULT 12
 QY 099054 PRELIMINARY: PRT: 6713 AA.
 ID 099054
 AC 099054
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE EBHA PROTEIN.
 GN EBHA OR SAI267.
 OS *Staphylococcus aureus* (strain N315).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; *Staphylococcus*.
 NCBI_TaxID=158879.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2131952; PubMed=11418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Ito T., Kanamori M.,
 RA Matsumaru H., Murayama A., Murakami H., Hoshiyama A., Mizutani-Ui Y.,
 RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
 RA Hiraoka H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
 RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
 RA Ogasawara N., Hayashi H., Hiramatsu K.,
 RT "Whole genome sequencing of methicillin-resistant *Staphylococcus*
 RT aureus".
 RL Lancet 357:1225-1240(2001).
 DR EMBL: AF003133; BAB42527.1;
 DR InterPro: IPR02988; GA.
 DR Pfam: PF01468; GA; 1.
 KW Complete proteome.
 SQ SEQUENCE 6713 AA; 722339 MW; AFEEDE226BE488 CRC64;

Query Match 5.8%; Score 622; DB 16; Length 6713;
 Best local similarity 20.5%; Pred. No. 4, 1e-12;
 Matches 559; Conservative 317; Mismatches 970; Indels 876; Gaps 124;
 QY 6 KVFENKATGTETMAVAVYAKASHSTGSCCATGQVGSYRTLSFRIALAVLVIGATLNGSA 65
 DB 39 ETLNKGOTGP--NTAKTAVDQALNNVNSAKHALNGTQNNKNAKQALITAINASDLNKGQ 96
 QY 66 YAGISIGSEADGGKGANAG-----DKSIAIGDIAQALGSGSIAIGDKKIYHNSNNAN 119
 DB 97 KDALK--AQANGAQRVSNADVDORNATELNTAMGOLQHAIDATDNTLASSKYVNAADSTRON 155

| | | | |
|----|------|---|-------|
| OY | 120 | IGKASGENSEJALGCGVNLASGHSJASIGSDDLUYLKEFTVQOISELLPIIRQKMLNIDY | 178 |
| Dd | 156 | AYTKRYTNAEHIJISGTEPVVTTTSEVYTAAN-----QVNSAKQELNDEKRL-- | 202 |
| OY | 179 | QLABTNLOKRYRTHAQGHASTAVGAMSYPKGSMAFGTRATAECTSYSLAVGLTATARA | 238 |
| Dd | 203 | -----VAKONAMTALDLO-----LMPQAK | 225 |
| OY | 229 | SSIAVGSMAQALIGRAATAVGSSTOVNLNKGIALFGSOYLQKDNDVYNAANPAY-----A | 293 |
| Dd | 226 | LKEQVQANRLIEDVQSVQTNQGSILNNAKGL-----RDSIANETTVKASQNYTDA | 275 |
| OY | 294 | PDDNOPIDNRKATKEKFGANDVFESIGNSGNDISIRKIIINVGSAGSDPAVVAO--LKE | 351 |
| Dd | 276 | SPNMO-----STYNSAVSNAGIINQTNPT-----MDTSIITQATQVYNAKNGING | 323 |
| OY | 352 | AVRLANROITEK-----GDDSNKRVKGLKTLITIGAQOT | 387 |
| Dd | 324 | AENLRNQNTAKQULNLTLSHLTNNOKSALISSQIDRAGHVSEYTAAKNATTELNAQMGILE | 383 |
| OY | 368 | SALIDHNIQVONG-----DGLK-----VOLAEFL-----TSIKVYTTENLTAN | 426 |
| Dd | 384 | QAIHDON--FTVKQGVNFTDADAKKRAYTNAVSRAETIILNKTOGANTSKQDVE---AAI | 437 |
| OY | 427 | EKVYVQKRLTTDKIGFTNMGIDESPLYDKDPIHAGGOKIRKLTAGVYDDDAAYG | 486 |
| Dd | 438 | QNTVSAKNALGDO-----NTYNAKNAKNAIINLTJISIN--NAOK--RDLYTKI--DDATVVA | 489 |
| OY | 487 | QLKRYVN-----QTAESALOTFTYVKYVD-----KNGDANDADSK-----IITYGKN--NKP | 528 |
| Dd | 490 | GVEAVSNMGTOLNMTAMNLNDGINDKANTLASBNHYDADSDKKTAYQAVTNAEMILKN | 549 |
| OY | 529 | DGTOVNTLKLGKENGVDVPTETNGIYVTFGLNONGNLTJVGNSTLNL--DGLSVKNTSNKQI | 587 |
| Dd | 550 | SGSNULD--KAABENALSOVTNAGALNGHNLEQAKSNANTTIINGLOHLLTQAOKDKLQO | 607 |
| OY | 588 | QVQADGITTFDIDSRSKGA--GIENTRIT--RGIGEFANNGSIDANKPRLPTIINAG | 644 |
| Dd | 608 | VQOQONAVGVDTPYKSSANTINGAMGTLRNSIDONTATKNGOYLDAT----- | 655 |
| OY | 645 | KELTNVOSAINP-----ATNGOLDPFRNLSTPANKESGSAATIKDYLNLSQV--PLTF | 696 |
| Dd | 656 | RNKNTNYNNAVDSANGVYNAISNPRMD--ANAINGIAFOVSTKNALDGHNLTOAQOTATN | 714 |
| OY | 697 | AGDTGPVNTKKGELIK-----VKGRTTADDLTKNNIGVADSTQNSLTVKLA | 745 |
| Dd | 715 | AIDATNTLNKQOKDALKAQVTSQORAVANTYSIQOTANEL--NTAMQOLQIGIDENAKOT | 773 |
| OY | 746 | KTISDDO-----AVNTKULIASDKTYVSG-----ANRKLONGLTF-- | 783 |
| Dd | 774 | QRYRDAQSKKTAAYDAVAAKAILNKQJGNSNDKAADVBRALOQVYSTRKDALNGSLAE | 833 |
| OY | 784 | -----SKONTGA-----TPATNSKITG-----VDGLKFTDNNGIALDGTVYITKQVGEA | 828 |
| Dd | 834 | AKAARONLGLNLHITTAQOTALEGOINOATYDVGVNTYKTAANLTLDGAMNSLQALN-- | 891 |
| OY | 829 | KODSGLDKSKPYLDKDKLKVGEYEITTINGINAGKAITGLSMLTLDATNATGHTYTOGI | 888 |
| Dd | 892 | DKDATL--RNGVYLDADASK-----RNAVTAQVTAEBGLNKQJGNTSKADVNA | 940 |
| OY | 889 | VDSIDTKIRASIGDVLAGNGLKNNGDADPFSTYDVFINGANATTAKYVYDYGASVYA | 948 |
| Dd | 941 | LNAVTRAKAA-----LNGAENLNR-----AKTSATNTINGLNLNLTQLODNLKHQVE | 987 |
| OY | 949 | YDVAVDGTTHLTLGADGNKNO-----IGVKUTTL--TKT-----DAKGKAIFNS | 991 |
| Dd | 988 | QAQVY-----GVNGVYKQKGTMLNTAMGALRTSLOQNDNTTKTSIONTILSDSKNNYN | 1040 |
| OY | 992 | VNSGDDKALINAKD-----IADNLTLAGEIRNTKGTADTALOTPOVKRYVKEGSDJDNAD | 1047 |
| Dd | 1041 | TAVNANAGVYNATNPRMDANALINDMANOVNTTKAALNQAOLQAQ----- | 10855 |
| OY | 1048 | TIYGRKAKTRNOVT-----LKLKGNGLDIOTNRKDGTVTFINGINTOSGLAAGNNTTLN | 1101 |

| | | | | | | |
|----|------|---|---|----------------------|----------|------|
| Db | 1086 | ----- | AKTNAITINNADDLNOKOKDALKTQVNNAGVSDANNVQH----- | TATELNG | 1133 | |
| Oy | 1102 | NGLSIK----- | NTAGNEOIQVGADGVKFAKVVNGVYGAG-IDGTTTRIT----- | | 1143 | |
| Db | 1134 | AMTALKAALADERTKASGN----- | YVNADEKROAYDSKVTNAEMENITNGTPATLTVDVN | | 1190 | |
| Oy | 1144 | -REIEFAGI-NG-SLDKSRPHLSK-- | DGIMAGGKITITIOSGETAONSNDVATGK | 1195 | | |
| Db | 1191 | SAASQVMAAKTALNGDNLLVFAKEHANNITIDL-- | AGLNVORAKIKEVOQSATYITLDG | 1246 | | |
| Oy | 1196 | IYDTKTELEENKISSTAKTQNSLHEFSVADE----- | OGNFTVSNP--SSYDTSKTS-- | 1246 | | |
| Db | 1247 | VQYKNS----- | SQTLTAKKGLRD-SIANEATIRAGQVYTTDASPNNNREIDSATLAKA | 1300 | | |
| Oy | 1247 | ----- | DVITPEAGENGITTVKNGGVAVGIDQTKGLTTPK----- | LTVGNNGK | 1289 | |
| Db | 1301 | IINQSNPTMEPTITQIA-TSQYTTKEHALNGAQNLAQK-- | TFAKNNLNLNLTSSINNAOK | 1357 | | |
| Oy | 1290 | GIVIDSQNGQNTITGLSNTLTANTYNDKGSVRTTEQCKI | IKKEDKTR----- | 1335 | | |
| Db | 1358 | DALTRNIDGATTVAGVQOETAKETELNMAHSLONG-- | INDETQKQOYQKYLDAPSKS | 1415 | | |
| Oy | 1336 | ----- | AASTVDSLSGFN-----LOG----- | NBE----- | AVDEPSTY | 1362 |
| Db | 1416 | AYDOVNAKAKIILTASGQVNDKAAVEQALQVNSITKITALNGDAKLNEAKAAKQITGLT | | 1475 | | |
| Oy | 1363 | DTVFADGNATTAFTAYD----- | DYSKTSKVYVD----- | VNVDDTYI----- | EV | 1401 |
| Db | 1476 | THINNAQRNLNDNEITQATVEGVNTVYKKAQOOLGAMQLETSIRKDDTLQSQNOQDA | | 1535 | | |
| Oy | 1402 | KDKKLQYKTTTLTSTGANKFKLSQATGDALVKASDIAVHALNTLSDDIOTAKASQAN | | 1461 | | |
| Db | 1536 | DDAKRTAYSAVNAAMAILNKTAGGNTPRAD-- | VERAMQAVTQANTALNGIONLERAKQAA | 1594 | | |
| Oy | 1462 | ----- | SSAGVYVADG----- | NKYIYDSTDKY | 1483 | |
| Db | 1555 | NTAITNASDLTKQKELKAOVTSAGRVSHANGVEHTATELTAMTALKRALADKADR- | | 1633 | | |
| Oy | 1484 | YQAKNDGTVDKTEVKADKLVLAAQO----- | TRDGTIAOMNYS----- | YIN-KEQVN-DAN | 1532 | |
| Db | 1654 | ASQNVYNAADANKRQARDEKVTAAEHIVSGTPTPLTPESDVYNTATQYTNKKTQOLNGSHN | | 1712 | | |
| Oy | 1533 | KKQGINEDNAFVGLERKAASDNKT----- | NAAVTUGDLNAVQPTPLFPAADGTGTAKKL | 1567 | | |
| Db | 1713 | LEVAKONANTALIDGLTSLNGROKAKLKEOVGOATLPLNQVTRDNAOTL-- | NTAMKGL | 1768 | | |
| Oy | 1568 | GETL----- | TIKGGQTDI-- | NKLTENNIGVVA----- | GTDEPTY | 1619 |
| Db | 1769 | RDSIANETITKAGQNTYDASQNKQNDYNNAAVTAKAIIIGOTTSPSMIAQETINQAKDOYTA | | 1828 | | |
| Oy | 1620 | KL----- | AKDLTNLSYSVA----- | GGTKIDD----- | 1640 | |
| Db | 1829 | KQOALNOENLRTAQOTAKOHLNGLSBDLTNAQKDAKQRIEATHVNEVTOAQONNADALN | | 1888 | | |
| Oy | 1641 | ----- | KGVSEYDSSGQAKANTPYLSANGDLDGKVIS----- | NVGKGT | 1678 | |
| Db | 1869 | TAMTNLKNGIODOQNTIKQGVNFTDAD-EAKRNA-- | YTNAAVTOABQILNKAQGPNTAKDG | 1944 | | |
| Oy | 1679 | KDTLAAVQOL----- | NEVRNLISLGNG-- | 1702 | | |
| Db | 1945 | VEPALQVNOARKNELNGONVWANAKTATTAKNALNLNLTSSINNAOKALKSQIEGATTVAGVN | | 2004 | | |
| Oy | 1703 | ----- | NDNAV----- | GNQVNIADIKKDPNSGS | 1724 | |
| Db | 2005 | QVSTMASELTAMSNLDORGINDPAALAKKAQOKTTEADROKQRAYNDAYVAAKTLLDRTKAGS | | 2064 | | |
| Oy | 1725 | SSNRFVTK----- | AGTVLGGKN----- | NDTEKLATGG | 1752 | |
| Db | 2065 | NDKNVAVEQALQVRNPTAKTALNGDALNEAKMTAKOQLATMSHLTNAQKAMLTQEIRGT | | 2124 | | |
| Oy | 1753 | VQVGV----- | DKDG----- | NANGDLSNVWYVTOQDGSKA | 1782 | |

Db 2125 TVAGVGIQANAGTINQANQNLQROSTASKDATKSSSEDDYQDANADLQNA-----ND----- 2175
Qy 1783 LLATYNAAGTNTLNNPAPALIDRINEGIRFHVNDGNOEPVYQGRNGIDSSASGKHSY 1842
Db 2176 --AVTNEGIIIS--ATNPENNPDTINOKASQV-----NSAKSALNDEKL 2217
Qy 1843 AIGFO--AKAGCEAAVAAGTGTQAGNOSIAIGDAAQAGDQSIAGTG-----NVVAG 1893
Db 2218 AAAGKOTAKSD-----IGRLTDLNNAO--RTAANAEVDQAPNLAAYTAARKKATSLITAMG 2270
Qy 1894 KHSGAIGDPSTVAKADNSYGVNNQFTDATQTDVFGVNNITVESNSVALGSN---SAI 1950
Db 2271 NLKHALAEKONTKRSVNY-----TDADQPRQOAYDTAVTQAEATITNAGNANETQV 2322
Qy 1951 SAGTHAGTQAK-----KSDGTAGTTTATGATGYKGA 1983
Db 2323 QALNLQNLQAKNDLNGDNKVAQAKETAKRALASYSNLNNAQSTAAQSDINATTVADVTA 2382
Qy 1984 GQ-----TAVGAVSYGASGAEIRIONVAAGEVSATSTDAVNGSOLYKATQGIANATNE 2036
Db 2383 AONTANELNTAMGOLONGIND-----QNTYKQOVNFTDADQ--GKKDAY--TNAVTNAGI 2434
Qy 2037 LDHRHONENKANAGISSAMAMASMPQAYIPGRSMVTG-----GIATH--NGQ--- 2082
Db 2435 LDKRANGONMTKAO--VEALNQVTTAKNALNGDANVRAQAKSDAKANGLITLILNNAQKOD 2492
Qy 2083 -----GAVAV--GLSKISDNQ 2097
Db 2493 LTSQIEGATTNGVNSVYTKAQ 2514

RESULT 13
Q931R6 ID Q931R6 PRELIMINARY; PRT: 6713 AA.
AC Q931R6;
DT 01-DEC-2001 (TREMUREL. 19, Created)
DT 01-DEC-2001 (TREMUREL. 19, Last sequence update)
DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)
DN HYPOTHEICAL PROTEIN EBHA.
GN EBHA OR SAVI434.
OS Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158878;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanemori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani U.I., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Kamehisa K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Hattori M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus".
RL Lancet 357:1225-1240(2001).
DR EMBL: AP003362; BAB57596.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 6713 AA; 722306 MW; BBOC0536AC341BF5 CRC64;

Query Match 5.88; Score 619; DB 16; Length 6713;
Best Local Similarity 20.5%; Pred. No. 5.2e-12;
Matches 559; Conservative 317; Mismatches 970; Indels 876; Gaps 124;

Qy 6 KVINKKATGTMAVAEYAKSHSTGGSCATGQVSVFTLSPARIALAVLIGITLNGSA 65
Db 39 ETIINKQTGP--NTAKTAEQALNNVNSAKRALGTONLNNAKQALTAITANGASDLQOKO 96
Qy 66 YAGIGISEADGSGKANARG-----DKSIAIGDIAQALSSQSIADGNKIVHNSNNAN 119
Db 97 KDAIK--AQANQAGVRSNANVQNRATELNTAMGOLQHAIDKTTLASSKYVANDSTKON 155

Qy 120 --IGAKASQNESTIAIGDVLASGHASIAIGSDLLYLKKTETVQOISELLPIRGOKALNDY 178
Db 156 AYTKVTAHEHITSGFTVYVTPSEVTAAN-----OVNSAKBELNDEKL--- 202
Qy 179 QIADTILQYRPTHAGHASTAVGAMSYAKGHFNAFGTRAEGETYSIAVGLTATARA 238
Db 203 -----VAKQANNTAIDALTO-----LNPQAK 225
Qy 239 SSTANGSNAQIGFATVAGSTQVNNLNGIALGFGSOYLQKNDVNAANRAY-----A 293
Db 226 LKEOVQANRLIEDVOSQJNGOGLNNAMKGL-----RDSIANETVYKASQNTYDA 275
Qy 294 PDONCPIDNRKATFNGKATDVFSIGNSGNSIRKIIINVGAGSADDAVVAQ--LKE 351
Db 276 SPNNQ-----STYNSAVSNAKGIINQTNPT-----MDISATIQATVANNANGLNG 323
Qy 352 AVLANRQITPK-----GDSNNRVEKGLKTLITGAQT 387
Db 324 AENLRQAONTAQNLTLSHLTNQKSAISSQIDRAGHSEVTAKKNNATELNAQMGNE 383
Qy 388 SALTDHNTIGVONG-----DGLK-----VOLAEFL-----TSIKVYTNELPAN 426
Db 384 QAIHDON--TVKQGVFTDADKAKRDATYNAVSRAETIILKTOGANTSQDVE---AAI 437
Qy 427 EKYTVKTRILTDKIGFTDMNGIDESKPYLDKDTGCIHAGGCKITRTAGVYDDDAATYG 486
Db 438 QWVTSKNAALNGDQ-----VYTNKNAKNAKNAALNNLSIN--NAOK--RDLTGTI--DDATVYA 489
Qy 487 QLKRVN-----QTASALQTFYKRVK-----KNGDANDSK-----IITVGN--MKP 528
Db 490 GVEAVSNTGTQNTAMANLQNGINDKANLASENHADSDKRTAYQAVTAENILKN 549
Qy 529 DGTQVNTLKGENVYDTEETNGVTFGLNQNGILVNSSTLNN--DGLSVKNTSNKQI 587
Db 550 SCSNLD--KAAVENALSOYTNNAKGAALNGNHNLEQAKSNANTTINGIHLTQAOKDKLQO 607
Qy 588 QVQADQITTTDISNSKPGA--GIENTRIT--RDGIGFANNTSIDANKPRLPTGIMNG 644
Db 608 VQQAQVAVGVVYKSSANTLNGMGTLRNSIQDNATKNGVYLATE----- 655
Qy 645 KELLTNOASINP-----ATNGQDLDFMRLSTANTEKSGSATIKDLYNLQV---PLTF 696
Db 656 RKNNTNNAVDSANGYINATSNPMD--ANAIQALQVYSTKNALDGTNHLQAOQTAN 714
Qy 697 AGDTGPNVTKIGELIK-----VKGKTTADILTKNNIGVADSTDNLSLVKLA 745
Db 715 AIDGATNLNAKQADLAKQAVTSQVRVANTSIQOTANEL--NTAMGOLQHIDDENATKOT 773
Qy 746 KTLSDLD-----AVNTKTLTASDKVTYDSC-----NNTAKLQNGDLTF-- 783
Db 774 QKYRDEQSKRTAYDQAVAAAKILNKQGSNDKRAVRALOQVYSTIKDALNGDKLAE 833
Qy 784 -----SKONTGA-----TPATNSKITG-----VDGLKFPDNGIALDGTYYTKRGVGA 828
Db 834 AKAAQONLGLNLTNAQRTALEGOINQATYDGVNTYKTNANTLTDGAMNSLQGAIN-- 891
Qy 829 KQDSLDKSKPYLDKDKLKVGEVEITTINGINAGKAITGLSNTLTDATNATGHTYQIGI 888
Db 892 DKQATL--RNQNTYLDADSK-----RNAYQAVVTAEGILNKQGTGNTSKADVNA 940
Qy 889 VDSIDKTRASTIDVYNAGENLKNNGDAADPVSTYVTFPIGNATTAKVYTDGASKA 948
Db 941 LNVATRAKAA-----LNGAENLNN-----AKTSATVINGLPLNQLQDQNLKHOE 987
Qy 949 YDVNVGQTTIHLGADGNKO-----IGVKTTPL-----TKT-----DAKGDAINFS 991
Db 988 QAOQV-----VGAVGVDKQNTLNTAMGALRTSIQDNNTKTSQNTYLDADSKNNYN 1040
Qy 992 VNSGDDKALINAKD-----IADNLTLAGIIRNTKGTADTALQTFQVKKYKENGDDNDAD 1047
Db 1041 TAVNANNGVYINATNNNMAMALINDMANOVNTTKALNGAQNLIQ----- 1085

| | | | |
|----|------|--|--------|
| QY | 1048 | TTTGGKADKCTQVNT-----LKLKKGNGSDIDTINKDGVYTGKINQSSKLKAGNNPTLNN | 1101 |
| Db | 1086 | -----AKTATATPTINNADDLNOKOKDALKTQYVNAQRYSDANNVQH-----TATELNG | 113338 |
| QY | 1102 | NGLSIK-----NTAGNEQIOVGADGVKFAKVNNGVVGAG--IDGTRTIT----- | 11433 |
| Db | 1134 | AMTALKAALADKERTKASGN--YVNAQDCKRQAYDSKYTNAMENIINGTRPATLTIVNDVN | 11909 |
| QY | 1144 | --RDEIGFAGT--NG--SLDKSRPHLSK--DGINAGCKKITNIQSGELIAQNSNDAYGSK | 11995 |
| Db | 1191 | SAASQVNAKATALMGDNNLNREKAEKHNATIDGL---AOLNNVQAKLKEQVQSATITLDG | 12466 |
| QY | 1196 | IYDTLEKTELEKISSTAKTQANSIHEESVANDE-----QGNFTYSNP--YSSYTSKTS--- | 12466 |
| Db | 1247 | VQYVKNs-----SOTLTAKKGLRD-SIAEATITKAGONTTAPSPPNREYDSAYTAKA | 13000 |
| QY | 1247 | -----DYTFPAGENGITTYKNGGVVAVGIDQTKGLTTPK-----LTVCNNNGK | 12898 |
| Db | 1301 | IINQTSNPTEMRPTITQA-TSQYTTKEHALNGQNLQAQK--TTAKNNLNINLTISINNAQK | 13570 |
| QY | 1290 | GIVDSQNGQNTITGLSNTLANVTNDKGSVYRTTEQSKIIKDEKTR----- | 13355 |
| Db | 1358 | DALTRNIDGATTVAGVQDETAKATELNMAHSLONG--INDEYQKQTKYLDABPSKKS | 14151 |
| QY | 1336 | -----AASIVDLSAGFN-----LQG-----NGE-----ADVPSTY | 13669 |
| Db | 1416 | AYDOAVNAKAILTRKASGVNDKAAREQALQNVNSTRKALNDBADAKLENAKAAKQTIGLT | 14757 |
| QY | 1363 | DTVFEADGNATTAFAVTVD-----DTSKTSKVYVD-----VWDDTPTI-----EV | 14015 |
| Db | 1416 | THINNAQGNALNDIITQATVBEGVNTYKAAQOOLDBAMQOLETSIRKDDTITQOSQNYODA | 15353 |
| QY | 1402 | KDKILGVKTTTTLTSTGANKFALSNQATGDALVYKASDIAVHLNLTSGDIOYAKASQAN | 14611 |
| Db | 1536 | DDAKRTAYSAQVNAATAATLTKTKTAGGNTPKAD-VERAMAQVTOANTALNGIQMLERAKQOA | 15944 |
| QY | 1462 | -----SSAGVYVADG-----NKVYISDIDMKY | 14833 |
| Db | 1355 | NTAITNASDLTKOKALKQAQVTSAGRYSAANGVEHTATELNTAMTALKRALADADAPK- | 16533 |
| QY | 1484 | YQAKNDGTVDTKTEFVADKDLVLAQAO-----TPDGTILAQNNVKS-----VIN-KEYVN-DAN | 15323 |
| Db | 1654 | -ASQNYVNAADANKRQAQADEKVTAAEHVYSTPTPTLTPSDVYINNAQTQYTNAKTQUNGNHN | 17122 |
| QY | 1533 | KKGINEDNAFVKGLEKSAASDNKTK-----NAAVTGDLNAAVQATPLTFPAGDTGTAKKL | 1587 |
| Db | 1713 | LEVAKONANATIDELTSLNGPQAKLKEQVGAQTTLTPVQTVRDNAQTL-----NTAMKGL | 17688 |
| QY | 1568 | GETL-----TIKGOTDP-----NKLDTNNNGVVA-----GTDGTV | 16191 |
| Db | 1769 | RDSIANEATITKAGONYTYDASQNKQNDYNNAAVYTAKAIIIGOTSSPSMAIOELINQAKDOYTA | 18238 |
| QY | 1620 | KL-----AKDLTNILNSVNA-----GGTKLID----- | 16400 |
| Db | 1829 | KQOALNQGQENITRQATPAKHGLNGLSLDITNAQKDAKAKROIIEATIHVNEVTEYQAQNNADALN | 18888 |
| QY | 1641 | -----KGVSFVDSGGAQKAMTPLYLSANGLIDLGKVIS-----NVGKGT | 1678 |
| Db | 1889 | TAMTNLKNGIQDQNTIKQGVNFTPDAD-EAKRNA--YTNAAVQABQILNKAGQPYTAKDG | 19444 |
| QY | 1679 | KDTLAAVQOOL-----NEVRNLIGLGNAG----- | 17022 |
| Db | 1945 | VETALQVQVQRAKKELNGNOMVNAKTTAKRNALNLTISINNAQKAALKISOIEGATTVAGVN | 20044 |
| QY | 1703 | -----NDNAD-----GAGVNIADIKKQPNSGS | 17244 |
| Db | 2005 | QVSTMASELVTMNSNLORGINDLAAATKAAQKTEADRDQOTLYNDAVYTAATKLIDKTAAGS | 20644 |
| QY | 1725 | SSNFTVTK-----AGTVLGSGKN-----NDTEKLATG | 17522 |
| Db | 2065 | NDNKVAVEQALQVNPFAKTALNGBADRLNEAKNTAKQALATMSHLTNAQKANLTBQIEKGT | 21244 |
| QY | 1753 | VQVGV-----DKDG-----NANGDISNVWVTKQDGSKKA | 17822 |

| | | | | |
|----|------|--|-----------------------------------|------|
| Db | 2125 | TVAGVQGIQANAGLNLQANMOLOKSTASKDATKSSDYODANADQNNY | ND | 2175 |
| QY | 1753 | GLATYNAAGOGTNTYLTNNPAPAIIDRIEBOGIRFEHHVDNGQEEVVOGRNGIOSSASGKHSV | 1842 | |
| Db | 2176 | -AATNNEGIIIS-ATNNEPNNPTIINQKKSQV | NSASALNGDEKL | 2217 |
| QY | 1843 | AIIGFO-AKADGEAAVVAIGROTKQGNOSIAIGDNAQTGDQSTAIETG | NYVAG | 1893 |
| Db | 2218 | AAVQOTAKSD-IGRLTLDLNNQO-RTAAAEVDAQPNLAAATKAKKATSLNTAMG | 2270 | |
| QY | 1894 | KHSGAIGDPSTVKADNSYSVGNNOPTLDAQTDQDFEVGANNITVESSNVAIAGSN | SAI | 1950 |
| Db | 2271 | NKTHALEKDKNTKTSVNV | TDADQPKQOAYDTAVIQAALITNAGSNANETOV | 2322 |
| QY | 1951 | SAGTHAGTOAK | KSDOTAGTTTAAAGTGVKGA | 1983 |
| Db | 2323 | QALNLQANLNDINGCNKTAQAKETRAKRAKLAYSISLUNNAQSTAAITSOIDNATTVADYTA | 2382 | |
| QY | 1984 | GO-TAVGAVSYVGASGAEERRIQONVAAGEVSATSDAVANGSOLYKATQGIAMATNE | 2036 | |
| Db | 2383 | AQNTANLEMTAMGLOQIND | QNVYKQOVNPTDADQ-GKKDAY | 2434 |
| QY | 2037 | LDHRIHQENKANAGISSAMAMSMQOATIPGRHSYTG | GIATH-NGQ | 2082 |
| Db | 2435 | LDKANQONMTRKAO-VEAALNQVTTAKNALNDGANDANYARQAKSDAKANGLITLHLLNNAQOD | 2492 | |
| QY | 2083 | -----GAVAV-GISKSLDNGO | 2097 | |
| Db | 2493 | LTSOLEGATTVYNGVNSVKYTKAO | 2514 | |

| RESULT | 14 | | |
|--------|--|--------------|---------------|
| 093DC7 | | | |
| ID | 093DC7 | PRELIMINARY; | PRT; 2065 AA. |
| AC | 093DC7; | | |
| DT | 01-DEC-2001 (TREMBLrel. 19, Created) | | |
| DI | 01-DEC-2001 (TREMBLrel. 19, Last sequence update) | | |
| DR | 01-DEC-2001 (TREMBLrel. 19, Last annotation update) | | |
| DE | RSCA. | | |
| CN | RSCA. | | |
| OS | Yersinia enterocolitica. | | |
| OC | Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; | | |
| OC | Yersinia. | | |
| OX | NCBI_TaxID=630; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=21437624; Pubmed=11553561; | | |
| RA | Nelson K.M., Young G.M., Miller V.L.; | | |
| RT | "Identification of a Locus Involved in Systemic Dissemination of | | |
| RT | Yersinia enterocolitica."; | | |
| RL | Infect. Immun. 69:6201-6208(2001). | | |
| DR | EMBL:AF394927. AA:77860.1; ; | | |
| SO | SEQUENCE 2065 AA; 210024 MW; A306D0763682D176 CRC64; | | |

[illegible]

Db 159 LFGKQAVDVGGLIATSTSMNOPISGCYLTLS-----OKQEGKLVNOANLFTTA-GGYI 213
 Qy 227 LAVGLATFAKAASIAVGSNAQAIPEATPAVGSTOVNLRGIALGF--GSOY---LOK 280
 Db 214 ALIGQOVNOBPQGVINTPO-----GKVALASSRYILNDRGNLLGVYOQGVQVTLLOIN 268
 Qy 281 DN---DYNAANVAVAPD-----DNQPIDNRKATFKNGATDVFSIGNSGNDI-- 327
 Db 269 GGLIRADEGVIGLTAGQKEMLMNTVIDNTGILQARGLSKKNV-----IYLNGBEGVYS 323
 Qy 328 RKKIINV---GAGSADDAVVAOLKEAVRL-ANRQITFKGDDSNRY----- 371
 Db 324 QQGMINVNSQQRG-----NVILIGENIHLVAASKIDARDEGEGKVLVGDWQGNK 377
 Qy 372 -----EKGKLTLLT-----TG-----GAQTSAL 390
 Db 378 LTKNARSVMOKGANIDVSSHQAGTAIVLSEHYTGEGIDIHARGSLSGDGGQVETS 437
 Qy 391 TDHN-----IGVONG--DGKLVOL-----AETLT 413
 Db 438 SQRNLOSGFGRVDVSAIMGNTGRMLDPAEVNTVSGAESGVSVOIGDIPAGYKKAQVFT 497
 Qy 414 SLKMYT-TEMLTANKVYV--KTLTITDKIGFTN-----DMNGI--DESKPYLDK 459
 Db 498 PMAVYTOILNTSINAOBKGTNVITTSNGSLTNCRCMNITLQADINKIAGDATTLOA 557
 Qy 460 DTGHAAGOKITKLTAAGVDDAATYTGOLKKVNOFAESALQF-FYKRVYD----- 508
 Db 556 DQNIYSNGHNT-----ATTGKLNLLSGDSITVSDITLLNNSDVLNCGDILL 606
 Qy 509 KNGDANDSKITTVGNKPKDGTUVNTLKGKENGVDVTTETNGVTEGLNONGNLVGN 568
 Db 607 KHANENMARISIMGR-----YQVGNLTLEBNTGM-----ASYGVNISNANI-- 651
 Qy 569 STLANDGLSVKATNSNKO-----IQVADGI-----TFPTDISNKPQ-----AG 607
 Db 652 SVAGETRISGESSNANOQWGRGIDISGDSVPAGKGNMSTFTMSNRSMMGFTNATJAG 711
 Qy 608 IENTT-----RITRDGIGFANNITSL-----DANKPRLPPT--GINAGKELT-- 648
 Db 712 DKNITFOANANGSTSGVDFTN--GSLVSKSGNISFEDINGETITOTSGLRIOGSGVN 769
 Qy 649 NVQSAIN-PATNGQLDFPMNLSTANTEKSGSAATIKDLYMLSOVPLTFAGDTGPNYTK 707
 Db 770 NVNVEINTKGVYDGLRBSHITATAGNISANATTHKGLM-----ISGDTLMSKN 821
 Qy 768 LGEILKVGKGTADDLTKNNIGVADSTDSLVYKLAKTISDLDAVTKTLTJASDKYTV 767
 Db 822 ---IKLOGVTNSTYVGAADAIKISGNSSSVOVNMAAGNIS--MVAVKNKTEVEGS---TI 873
 Qy 768 DSGNNTAKLONDLEFSKONTGAPATNSKITVYDGL-----KFTDNNOIALDGTIYIKD 823
 Db 874 SADYINIAAQSGDFENLISGVKGFPSN-VNISADNTILNNSIDNDVAVNTNFFLA--- 929
 Qy 824 KVGFAPKODSLDKSP-----YLDKDLKVG--VEITNGINAGKATLTGSLNLTLD 874
 Db 930 ---AKGDIKANLSSPNKALYFKNGSGMSAGONSIVAVNVSAGSVEAVTITGLA---SN 982
 Qy 875 ATNATTHVTOIGYVSDTKTRASIGCVLNAENLKNNGDAKDFVSTYDTVDFINGNAT 934
 Db 983 RNNVTVGR--DISIT--ANNHGMGTGAGIG-----VDYVNEFAK 1017
 Qy 935 TAKVTVYDGKASVAVADVNDGTIHLTGADGNKNOIGVKTTLTKTDAKGKALNFVSN 994
 Db 1018 NGNFTANNNGSK-----SIGIANADIRAEVNLIISNT--SRADSVIIRANITTLT 1066
 Qy 995 GDDKALINAKDIADNL-----NTLAGE--IRNTKGTADTALQTFQVKKVKEGDD 1043
 Db 1067 GN---IANANTSTNRGIVIRONTLSAQKELMLAATSSASPAI-----IYOGLSDE 1115
 Qy 1044 NDADITTYGK-----DAKTQOVN-TLKLKKGKGLDIQTNK--DGIVTFG-- 1084

Db 1116 SRSHLVAQGNISLKGNSKSGNPRSSVNLANSVLTSGKN-IDINSSVGDGDYFENNVD 1174
 Qy 1085 INFOGLKAGNNTLLNNGNISIKNTAGNEQIOVA-----DOVKRAKYNNGVANGI- 1136
 Db 1175 LNALG-----NVTVYGBALSAISTATNSVLTSGNNSIKALNGLLIGKALINTSGAGTL 1229
 Qy 1137 -----DGTTRITRDEIGFAGNGLSLDKSPHLSKQD-----INAGG 1172
 Db 1230 FRANGSLSVAGNIALQGETGTGAIRNGIAFYGANNTLINAKDSLSLIGENTSQITAGG 1289
 Qy 1173 KKITNIOGELIAQNSNDVATGKITYDKTELENKISSTAKTAONSHEFSVADQGNFT 1232
 Db 1290 NGISVLSPKTTLINNNSLT-----MEGRSTGA-----GINFP 1323
 Qy 1233 VSNPYSSVDYNTSKTSDVITFAGENGITK---VAKGVYPRG-IDQTKGLTTPKLTVGN 1287
 Db 1324 IGN-----NTVLNGBEGSLIKSGSVAGSVAISGMVNSSGPAT--IEGSTD 1370
 Qy 1288 GKGIIVDSQNGQNTITGLSLNTLANVNTDKGSVPTTEGKIITKDEKTRAISIVDSAGF 1347
 Db 1371 GSGVHLFSAEHQ-----IKRINVTGSSIOABGLRI-----SGNATITDTALSGK 1414
 Qy 1348 NLQNGEAVDVFSTYDVTNADGNATTAKVYDDFTSKSVYVDVNDVDTTIEVKOKKLG 1407
 Db 1415 SINGSGIKVDSLT-----GSGVITHVLDMAALNGSSTHIGVE----- 1453
 Qy 1408 VKTITLSTGTGAKKFAKLSNATGDALVYKASDIYAHN---TSGDLOTKAKGASQANSSA 1464
 Db 1454 ---ITSDINGIHOSITIN--GTTDGIYGDIDIDKLNHVTGETSETDILLQGVATTGSGT 1506
 Qy 1465 GYVDADGNKVLYDSTDKKYYQAKNDGVDPDKTEVAKKDLVAQAQOT-----PDGTLAQ 1517
 Db 1507 GIKLDDNDLNSN-----SLNGAIVDGIALLDISLINSNGTRINGRASDGIGYQI 1558
 Qy 1518 N---VKSVINKEOVNDANKKOGIMEDNAEYKLEKAASDNKTKNAAYVGDINAQOTPL 1574
 Db 1559 NGTSLNGVNGTSSANGVGVQVSGSLDSDSHVTGISASSGVYD-----AETVL 1606
 Qy 1575 TFAADTGTITAKKIGETLT-----IKGGOTD-INKLTLDNNIGVAGTDEFTYKLA 1623
 Db 1607 NNTMLKGTSTHDGIGVETITANLSGYGSSVOGDVDTGTVBRIDRNTLLVYG----- 1656
 Qy 1624 DLTNLNSVNAAGTKIDDKGFSVDSGQAKANTPVLSAN-----GLDLAGKVISNYNGKT 1678
 Db 1657 DTDDLVIYNMGATODQGRGLOL--NGNNTLLDNTTLAGNMSDGTGIDIDP-LTKNGNST 1712
 Qy 1679 KD---TDAANYQOLNEVRNLLGNAAGNDNADGNQVNIADIKDP---NSGSSSNFTYIK 1732
 Db 1713 VDGKATDGDGVQLNGAISG--GYVNGTSDTSGIKVD--GDSELDNATILNNGNSPDGKIEI 1769
 Qy 1733 AGTVLGGKGNNDTEKLATGCVQVGVKDKGNANGDLNWNVWKTKOKDSKRALA-TYNAAG 1791
 Db 1770 VANLTGNGHSAVHETAEGS--GVDIGONA-----TLTGGTNDLLAVTQNASG 1816
 Qy 1792 QTNVLTNNPAAIDRINEQIRFEHVNDGNOEPVQGRNGIDSSASGKSHVALIGFOAKAD 1851
 Db 1817 D-----VGTGYO--LDGNNTLLDNTTLTAGNA-----SD 1841
 Qy 1852 GEAANAIGROTQAGNOSIAIGDNAQATGDSIALGTGNVVAAGKHSAGIGDPSTYK----- 1906
 Db 1842 GHGIEVGVPSSTGNNTI--NGNTVGD-----GCVHIDGPMGSGLVNGNSANNHGIY 1892
 Qy 1907 ---ADNSYSVGNNOFTATQTDVF--GVGNNTIYTESNSVALGSNSAISAGTIAGQ 1959
 Db 1893 LNDYAAINDITLGNAGSGKSLMEIYLPKNIGGNVTI--NGRPIDKNSVCGRTKSGSTL 1949
 Qy 1960 AKKSDGAGTITTAGAGTVYKFAQTAVGAVSVGASGARIRIONTAAAGSVATSDAVN 2019
 Db 1950 ISPPASTSAPTPSPVPTPL-----LTPILISG-----ENTISQITQPOEISKH 1995
 Qy 2020 GSOIYKATQGIANTNELHRI-----HONENKANANGISSAMAMASMPQAVYIPGRSMVT 2073
 Db 1996 GSIMMKRNOILSS---LDEOILPPLVAVTESERDIANIS--VVVICPE---GETTES 2044

OY 2074 GGIATF 2079
Db 2045 GPCETH 2050

RESULT 15

Q98LN6 PRELIMINARY; PRT: 3659 AA.
ID Q98LN6
AC Q98LN6
DT 01-OCT-2001 (Tremblrel, 18, Created)
DT 01-OCT-2001 (Tremblrel, 18, Last sequence update)
DT 01-OCT-2001 (Tremblrel, 18, Last annotation update)
DE ML0950 PROTEIN.
GN ML0950.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099; PubMed=11214966;
RX MEDLINE=21082930; Pubmed=11214966;
RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Matenabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Yamada S., Nakazaki N., Shimo S., Sugimoto M.,
RT Takouchi C., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
DR EMBL, AP002996; BAB48427.1; -
KW Complete proteome.
SQ SEQUENCE 3659 AA; 337498 MW; 7B7BC5C0F3816 CRC64;

Query Match

Best Local Similarity 21.5%; Score 600.5; DB 16; Length 3659;
Matches 524; Conservative 234; Mismatches 855; Indels 821; Gaps 115;

OY 29 GGGSCATGCV---GSVRLSFARIALAVLIGATLNGSAY-----AGIGISEADGG 77
Db 1101 GGGSAKTAVTSTGDTITTYGDAQIGLAQSVGGGNGSTVSLALAKAGIGV--ALGG 1158
OY 78 KGSANRG-DKS-IAIGDIAQALG-----SOSIALGDKMIVNSNNNNI 120
Db 1159 KGGAAAGNGLDVYVISTGISTGAGTSSGRTGAAGLAAQSVGG-----GGNGGF 1210
OY 121 GAKASGNEISAIAGDVLASGHASIAIGSDLYLKETVQOISELLPIIRGKALNDIYOL 180
Db 1211 AGTLGGGKSTAVG---VAFGGSGAGGGSADI-VKVTSTGNIS-----1248
OY 181 ADTNLOKRYRTTAQGHASTAVGAMSTAKGHFSNAFSTRATA-----ECTYSLANG--- 230
Db 1249 ---TNPDN-----SSGIIAQSIGGGGNGGFSVAVTGALGPDATASVAIGKGG 1296
OY 231 -----LEATAKAASSIAVGSNMAQIGFAFPAVGGSTQVULNKGIALGFS 275
Db 1297 VGGVGADVTVTSTGTTTGGKFSNGILASIGGGGNGGFAVAGSATTG-MAGIGVGVG 1355
OY 276 QVLOKDNVNAANVRAYAPDNDQPIDNRYKATPFKNKATDVFSIGNSNGDSIRKKLIIN-- 333
Db 1356 -----TGAT-----GSASG-----KVIIVNSY 1371
OY 334 --VAGASADTDVAVNAOLKEAVRLANKQTTFKDDSNMREYKGLGTLITTTGAQTSALT 391
Db 1372 SLVGAANG--QVPLIAPASNIY-----SLWTOGDNSSGIFAQSVG-----GGGGGGGGA 1417
OY 392 DHNIGVQANDGLAKVOLAELTILSLKWTTPNLTANKEVYTKQRRLTDTKIGFTNDNGID 451
Db 1418 G-SLGVOLGGGGLVSGGAGS-----GSTAETVYVTSNNIITGGK 1459
OY 452 ESKPYLDKDTGTHAGGOKITKLTAGVVDADATYGLKKVQTAESALQTFIVKVKDKNG 511

Db 1460 DSEFMAQSVGGGGGNGGFAALAGSKDMAASV-----AVGSG 1498
OY 512 NPDANDSKILTVGK--NNKPDGTOVNTLKLK-----GNG-----VDVETETNGVTFGLN 559
Db 1499 SSGGDAVAVTVVSHGNIETDDGSHAFPAQSVGGGGNGGAGLIGLIEGGISLAVG 1558
OY 560 QNNGLTVGNSTLNNNDGLSVKNTNSNKOIOVGADITFTDISNKPAGLIENTRTTRDGI 619
Db 1559 GS-----GONGNAKAVTVYSTGDLRTGELKADGILASIGGGGNGGFSGALATV---- 1609
OY 620 GFANNTGSLDANKPRLTPTGAINNGKELTNVQSAINPAINGGOLDPMNLSTAN--TEKS 677
Db 1610 -----TGAGAT-----GVVGGG-----GKGGNADTVYVTSIGNIYTLKN 1645
OY 678 GSAATIDLYNLSOVPITFAGDTGPNVT-----KKLGEILKY-----KGGKTADDTKN 727
Db 1646 GSNIGI-----LAQSVGGGGNGGAAVTTIAGSGKMAAAVSVGGGAKGG--TSQLVTVN 1697
OY 728 NIGVADSTNSLTVKLAKTLSDLDVAVNTKTLTASDKVTVDS-----GNNTAKLNG 779
Db 1698 NIGTI-DTGDKNAGLILASIG--GGGCTGFPALISEMVDGAGAAVSVGGNGTGDDG 1754
OY 780 DLTSKONTATPATNSKTIGVDGKFTDNNGIALDGTTYTKDKVGFAPAKODSLDKSRP 839
Db 1755 GRVIVNSNIGTTLANNATLHTVG--GDSNG-----TFQSVGGGGGGGGS---- 1799
OY 840 YLDKDLKGVFEVETTN-----GINAGKRA-----ITGLSNTLTDATNATTHY 883
Db 1800 -----GAVSTGNMAKAIGVSVSGGAGSGDGKTIVTVSVNIIITGG--NGANGIL 1849
OY 884 TO-----LGIYDSTDKTRAASIGDVLNKNFNLKNNDKADVSTY 923
Db 1850 AQSIVGGGGNGGFPFAATIKGSLGKNPSKNGAVSVSLGGAGAGG-----GDAEKV--TV 1901
OY 924 DTVPFI-----NGNATTAKVYDGGKASVAY 949
Db 1902 DSTGIITSGKRAFGVLAQSVGGGGTGLSVAAALNLEGCMQITIAVVGEGGGGTG 1961
OY 950 DVNV--DOTTHTLHGADGNKNOIGVKTTLTKTDPAKDAKINFSVSGDKKALINAKDIA 1007
Db 1962 EVLLTRHGSTI-TTG-----DQSVGLFPAQSVGGSGNGGMAIS--GVINGTD-----AKTIS 2010
OY 1008 DNLMTLAGEIRN-TKGADT--ALQTF-----QVKVKEGNDDDNDADOTIYVGDART 1057
Db 2011 ASVCGFGAGTNAKAVYVDNTGALSTYGVSDAIOASIGGGGNGGMAVSAVIGSLGTG 2070
OY 1058 NOVMTLKLKGNGLDIDTNKDGTVTFGINQSGLKACNNNTLNNNG-----1103
Db 2071 TNFAGVTVGFGDAGFAGDVETVNNGLLOTGLLAAGOTTTNDGAYGIFAQSVGGGG 2130
OY 1104 -----LSIK--NTAG--NEOIOVGADGVKPAVNNGVG-----AGID-----G 1138
Db 2131 SGNNAITVGLKGNNGTQVNVSAVG-----GSAGNGTGNATVIRQYGGIETNGMG 2184
OY 1139 TTRITREIIGFAGNIGSLDKS-----KPLHSDGAINAGKKTITNQS--GELAQMSND 1189
Db 2185 AFGILASIGGGGGTGGFRANSISLQLAGKCTLPVCEPAGKPRMWNLOATVGGAGGTND 2244
OY 1190 AVT-----GKIYDILKTELENKISSTAKTQNSLHESVADQ--GNNTFVSHPYSYD-- 1241
Db 2245 AATVDVGN--YDFTTDBDK--SSGIVAOSIGGGGIGDGVAVGTGGLAIPYVPVPTL 2300
OY 1242 -----TSKTSPIVTFP-----GENGITTKVKNGVVRVIGIDQTKGLTTPKLTIVGNN 1287
Db 2301 LKPLGTSLSLTSGVVALIGNGAGGNGGTYIVANEGVITTHGKSPDIIHQSIGGGGD 2360
OY 1288 GKGIIVDSQNGNTITGL-----SNTLANYTNDKSVRTTEQKII-----1338
Db 2361 GG-----DGFAGALGVIGGCGKAAGNGSVTVNKGKANIISOAGAILIQGTGSPTPPAE 2416
OY 1329 ---KDEKTRASTVVDLSAG-----FNLQNGEYVDFVSTYDTVNFADGNAT 1373

Db 2417 GABSADEQCYSAIGFAQSVGGGTGGAGALLSLGSGKA--GGTGHVTVNNYGGILL 2475
QY 1374 TAKVTYDDTSKTSKVYVDVVDTTIEVKDKLGVKTTTLTSTGTANKFALSNQATGDA 1433
Db 2476 HA-----DDSVGIFAQSIGGGGAGGSLGISAIAG-- 2506
QY 1434 LVKASDIVAHLNLTLSGDIQTAKGASOANSAGYVDADGNKVYDSTDNKYQAKNDGTVD 1493
Db 2507 -----GSGGASGSGGVYVNNMT----- 2526
QY 1494 KTEVAKDKLVQAQTPDGTLLAQMNVKSVINKQVNDANKKQGINEDNAFVKLEKRAAD 1553
Db 2527 ETHGI--DSYAIOAQSVG----- 2547
QY 1554 NKTNAAVTVYDNLAAVQPTLFFAGDTGTAKKLGFTLTIKGGQDTPNKLTDNNIGVAG 1613
Db 2548 GGGKNGCIT--GSIPL-----ISIGAGSSG--IGGIVN---TNNSL--HTYG--AG 2592
QY 1614 TDGFTVLAADLTNLSVNAAGTPIIDKGVSPVDSGQA--KANTPVLANGLDLGGKV- 1670
Db 2593 ADGI-----NAQSIGGGGS--GGRAIGFAVGGKGGDVGNGTTAGSNG--GGGAVT 2640
QY 1671 ISVVGKGTCDTDAANYQL-----NEVRLLGLGNAGNDADGNQVN 1712
Db 2641 VNNNANGITTEIGAHGIFAQSVGAGGSGGAFGVSIYPVAIGVGGGSSGDDGVYT 2700
QY 1713 IADIKDPNGS-----SSNRVYKAGTVLGKGNNDTEKLT--GGVYGVKDGNGANG 1765
Db 2701 V-----NNHGDITTVSASSVALFESIGGGGTGAMSVASAPAPGATFTGIGDGGANG 2754
QY 1766 DLSNVVKTOKDSSKALLATYNAAGQTNVLTNPAEALDRINEGIRPFHNDGNQEPV 1825
Db 2755 KGGNVNVTNFSOG-----IHTQGEGST-----GMAQSVGGG----- 2788
QY 1826 VQGRNIDSSASGK---SVAIGFOAKADGEAAVAIGROTOAGNOSIAIGDNAQATGDOS 1882
Db 2789 --GAGGASYVSOGPPGLAIALGKGAAGDGIIV---TVINNGAM-----QLDGDNS 2836
QY 1883 IAGTGNVAVAGKHSG-----AIGDPTVKADNSYSVGNNOFTDATOTDVFVGNNTIV 1936
Db 2837 VAIFFAOSVGGGGSGGTATAAALGVPIIGD-----TGAT-----GKGDVTV 2880
QY 1937 TESNSVALGSNSAISAGTHAGTQAKSDGTAGTTTACATGTVKGFAGOTAVGAV----- 1991
Db 2881 TMTGQIRLTGNG--SVGIFA-----QSVGGGGVYTAGTGTIVQAVAGSGNGGVYTTNS 2933
QY 1992 -----SYGASGAERRIONVAAGEVSATSDAVNGSOLYKATOGIANTNEIDHRI 2041
Db 2934 NVAMLITGDNSVDFG-----OSIGG-----GGVGFSGNYL-----GLDQVQS----- 2973
QY 2042 HONENKANAGISSAMAMASMPQAYI--PGRSMTGIGATHNGGAVAV----- 2087
Db 2974 -----TSSKMAA--POGFMGSAGGGGTGATFTOTADLAVIGKNSFALMEQSA 3021
QY 2088 -GLSKLSDNGQWVFKINGSDPTQGHVGAAGAGF 2120
Db 3022 GGTGNTSDNGDINVTIASGVTTIGSGAGGIGY 3055

Search completed: July 30, 2002, 15:38:43
Job time: 408 sec

1